BEHAVIORAL RESEARCH IN ACCOUNTING
Volume 20, Number 1, 2008
pp. 131–145

The Effects of Decision Aid Design on the Information Search Strategies and Confirmation Bias of Tax Professionals

Patrick R. Wheeler
University of Missouri–Columbia

Vairam Arunachalam
University of Missouri–Columbia

ABSTRACT: We report the results of a study involving 142 tax professionals designed to investigate the effects of decision aid design on information search (i.e., tax research) and confirmation bias. Results indicate that the participants exhibited confirmation bias when conducting tax research for clients. That is, participants showed a tendency to preferentially select information in support of their earlier recommendations to the client, even when the recommendation disagreed with the client's subsequent tax position. Results also indicate that while some decision aid features can reduce confirmation bias during tax research, others do not and may even enhance this bias. Specifically, a justification requirement decision aid reduced confirmation bias in terms of both the number and perceived importance of selected confirmatory cases, whereas a factor evaluation checklist decision aid either increased the bias (i.e., increased the perceived importance of cases) or had no effect on the bias (i.e., no effect on the number of cases). We suggest several decision aid design features for reducing confirmation bias in tax research.

Keywords: confirmation bias; information search; tax research; decision aids.

Data Availability: Please contact the first author.

INTRODUCTION

Tax laws often require taxpayers to choose between alternative tax positions. Effective tax preparers help clients under these circumstances to diminish the resulting uncertainty by explaining the risks associated with each alternative (Beck et al. 1991). Failing to provide accurate risk assessments can be costly for clients and practitioners during tax planning because the risk assessments provided may influence the tax positions that clients choose to file, and contested tax filings may end up in court. Similarly, during subsequent tax compliance (i.e., after the client has adopted a tax position), faulty risk assessments may affect how well prepared clients are for defending their tax position in court. Thus, it is important that tax practitioners communicate to their clients unbiased judgments concerning the likelihood that any given tax position will be upheld in court, facilitating clients in making sound decisions. Yet research indicates that tax practitioners

The extensive assistance of PricewaterhouseCoopers and the Missouri State University School of Accountancy Advisory Board in the pilot testing and conduct of this study are gratefully acknowledged. The authors also thank Paul Steinbart and Dan Stone for their thoughtful comments on earlier versions of this manuscript.

frequently exhibit confirmation bias when doing tax research for clients, e.g., preferentially selecting evidence supporting their prior beliefs (Barrick et al. 2004; Cloyd and Spilker 1999, 2000; Cuccia and McGill 2000; Hatfield 2000, 2001; Johnson 1993). Biased evaluations of tax evidence may result in skewed assessments of tax issues being provided to clients, putting clients at risk.

Tax decision aids, which exist in a wide range of technologies (Masselli et al. 2002; Noga and Arnold 2002; Vertex 2005), have the potential to mitigate these decision-making problems in tax practitioners. The effectiveness of decision aids in this regard depends on properly designing them to help users offset cognitive weaknesses and exploit cognitive strengths (Kachelmeier and Messier 1990). However, the effects of the design of current tax decision aids on tax decision-making have not been systematically examined. Using decision aid design research to help fill this gap, we predict that decision aids that prompt tax practitioners to evaluate the degree to which evidence is supportive or non-supportive of a given tax position will mitigate confirmation bias by promoting balanced assessments of the evidence. Alternatively, we predict that decision aids that prompt users to assess pieces of evidence individually and without reference to their relative importance will be ineffective at debiasing confirmation bias because they provide opportunities for selectively focusing on and overweighting facts supporting users' initial beliefs.

We investigate these predictions using an experiment involving 142 tax professionals. As expected, our results indicate the significant presence of confirmation bias during evidence selection. Results also support most of our expectations concerning decision aid design. Requiring participants to discuss both supporting and non-supporting evidence reduced confirmation bias. Requiring participants to separately evaluate checklists of preselected factors either amplified or had no debiasing effect on confirmation bias. These results indicate that the design of tax decision aids can have significant practical implications concerning evidence selection during tax research.

This paper is organized as follows. In the next section, we review the relevant literature and develop hypotheses. A discussion of the research method follows. The final two sections report results and conclusions.

BACKGROUND AND HYPOTHESES

Confirmation Bias

Social psychology indicates robust tendencies toward preferentially selecting and overweighting evidence confirming beliefs, and under-selecting and underweighting disconfirming evidence, in nonexpert decision-makers when engaged in a variety of tasks (Eysenck and Keane 2000; Plous 1993). Research suggests that decision-makers in general try to account for new data in terms of functional relationships that they believe already exist (Kahneman et al. 1982). In Heider's (1958) study, participants built explanatory causal models from confirmatory evidence, largely ignoring equally salient disconfirmatory evidence. Subjects even continued to support their beliefs with justifications derived from data that they were informed, subsequent to constructing the justifications, were unreliable (Anderson et al. 1980; Ross et al. 1977). The literature suggests that the heuristics underlying such dysfunctional behavior are beneficial under many, if not most, conditions (Wheeler and Jones 2003). Unfortunately, they are frequently applied to situations in which they degrade performance because they are hard-wired into the decision-making processes and are thus used automatically without critical reflection.

Consistent with these general findings, research specifically on tax professionals has found confirmation bias in a variety of tax settings. Johnson (1993) studied the effects of

confirmation bias on the information evaluation phase of tax research. Cloyd and Spilker (1999, 2000) expanded the study of tax confirmation bias by examining differences in information search and evaluation in response to client preferences. Hatfield (2000, 2001) and Barrick et al. (2004) extended research on tax confirmation bias to include the roles of managers and supervisors. These last three studies found evidence of confirmation bias, but also found that such effects could be moderated (e.g., by accountability to supervisors). Similar to distinctions made in the current study, Kahle and White (2004) differentiated information selection bias based on client advocacy[1] from confirmation bias (i.e., seeking evidence supporting the tax practitioner's prior beliefs, regardless of the client's tax position), and found that the former bias, but not the latter, was present in participating tax professionals.

Thus, except for Kahle and White (2004), the literature consistently indicates the occurrence of confirmation bias in tax practitioners, although the bias may be moderated by other factors (Hatfield 2000). However, whether confirmation bias is equally prevalent during the different stages of tax work (e.g., planning and compliance) has not been investigated. One might expect confirmation bias to be less salient during tax compliance because the tax practitioner's focus should be on collecting evidence to support the client's position regardless of the tax practitioner's earlier recommendation during tax planning. However, as noted above, research indicates that the heuristics that cause confirmation bias are remarkably robust and tend to be applied in an automatic or unconscious manner, even to situations for which they are inappropriate. Also, having one's recommendation ignored by the client might cause concerns for reputation and self-image to exacerbate confirmation bias. Thus, there is uncertainty as to whether confirmation bias will occur in the results of our study, especially considering that the confirmation bias is measured during the compliance stage.

Tax Research Environment

Tax judgments, like other accounting judgments, tend to concern technical issues within the decision-makers' areas of occupational specialties for which professional education and training have been received. The tax researcher must be able to locate relevant authority for a specific issue of interest from the myriad of information contained in a working tax library (Marshall et al. 1992). Thus, the potential for information overload is a characteristic of the professional tax environment (Rose and Wolfe 2000). In this regard, tax researchers often use one or more paper and/or electronic tax services (Magro 1999). Such tax services are decision aids in that they provide guidance for locating specific tax authorities and may provide short digests or annotations for individual court cases and IRS rulings. Nevertheless, it is ultimately the tax researcher who must decide which authorities to investigate further and read in full.

Given the massive number of cases on given issues and limited time, tax researchers must be selective in deciding which authorities to read. Therein lies the potential for a selection preference that might negatively affect the overall quality of the tax research process. As noted by Cloyd and Spilker (1999), without an understanding of authorities both supporting and opposing the preferred (or expected) tax outcome, tax professionals

[1] Based on prior client advocacy research (e.g., Johnson 1993; Barrick et al. 2004), our tax research model initially included client advocacy as an independent variable. Analysis of our advocacy data (discussed in footnote 6) indicated no significant relationships to other model variables. We therefore removed client advocacy from the final version of our model.

may misjudge the likelihood that any given tax position would be sustained by the courts. Thus, tax professionals need to obtain a balanced view of the evidence, and not preferentially focus on cases supporting (i.e., were ruled in favor of) one tax position over another.

Tax professionals frequently assist clients with planning and compliance aspects of possible tax positions. In general, planning involves offering advice on the different tax positions among which a client must choose, whereas compliance involves assisting a client after the client has adopted a particular tax position. Both stages generally involve tax research. Thus, it is not unusual for a tax professional to recommend a tax position to the client during planning that the client does not subsequently adopt. Accordingly, there are two possible information search biases which tax professionals may exhibit during compliance phase tax research, i.e., after the client has adopted a tax position. The tax professional may preferentially select information in support of either the tax position recommended during planning or the tax position adopted by the client at the beginning of compliance. Hereafter, we will refer to the former as a confirmation bias and the latter as a positive search strategy. This use of the term confirmation bias is consistent with the literature, which generally uses it to refer to the preferential selection of evidence to support one's prior beliefs (e.g., Cloyd and Spilker 1999; Kahle and White 2004). Both methods of tax research involve the preferential (i.e., biased) selection of evidence in support of one tax position over another. In cases where the client adopts the tax professional's recommended position, the two biases are indistinguishable.

Decision Aids

Decision-making research indicates that human judgments are prone to errors and inconsistencies (Eysenck and Keane 2000; Tversky and Kahneman 1974). Because of the inverse relationship between judgment errors and decision quality, a useful route to improving judgment performance may be through decision aids that reduce systematic judgment errors. However, the literature indicates that some decision aids improve judgment, while others impair it (Arnold and Sutton 1998; Benbasat and Nault 1990; Rose 2002). Accordingly, decision aid design is critical in determining whether decision aid use is productive (Glover et al. 1997; Kachelmeier and Messier 1990; Silver 1990).

We examine decision aids designed either to debias the heuristics that cause systematic judgment errors or to reduce the cognitive effort requirements in order to avoid cognitive overload, which typically degrades decision-making. We test these two designs with decision aiding techniques commonly used: a justification requirement (i.e., the user explains his/her decision) and a factor evaluation checklist (i.e., the user separately assesses factors identified by the checklist as important) (Ashton 1992; Jones et al. 2001).

Studies involving both nonspecialized and business decision-making indicate that justification requirements can improve judgment quality (Ashton 1992; Johnson and Kaplan 1991; Jones et al. 2001; Kennedy 1993; Tetlock and Kim 1987). Ashton (1992) summarizes the research up to 1992 by noting that justification requirements typically reduce the impact of information processing biases (e.g., overconfidence, order effects, and insensitivity to new information) while enhancing awareness of the potential beneficial uses of information provided by the decision aids. However, there is some evidence of justification impairing performance by causing experts to align their judgments to the preferences of others (e.g., Peecher 1996).

Research findings on the merits of checklists and other restrictive mechanical aids, albeit used in various forms in accounting, are more ambivalent than those on justification requirements (Ashton 1992; Bonner et al. 1996; Kachelmeier and Messier 1990). While

checklists can make decision-making more efficient, they frequently focus the decision-maker's attention on a narrow set of information that can accentuate any initial tendency to seek confirming evidence related to the mechanically selected factors.

Based on the research on these two types of decision aids, we expect that a factor evaluation checklist is more likely to support any existing confirmation bias, thereby impairing tax decision-making, than would a justification requirement decision aid. The focus of a checklist aid is on efficiently assessing factors (i.e., pieces of evidence) individually, rather than on a balanced evaluation and weighting of factors, as in the justification condition. A checklist, unlike a justification requirement, does not typically require users to consider the relationships among the factors, e.g., that evidence might be evaluated individually as highly supportive of a tax position, but assigned little value relative to other pieces of evidence and thus be of little overall importance during decision-making. Also, a checklist aid does not usually address or elicit other dimensions or facts of the case, such as those that might have come up in a broader assessment of the situation when writing a justification requirement. Accordingly, we hypothesize:

H1a: Compared to subjects without tax research decision aids, subjects provided justification requirement decision aids will select fewer tax cases supporting their recommendations.

H1b: Compared to subjects without tax research decision aids, subjects provided factor evaluation checklist decision aids will select more tax cases supporting their recommendations.

H2a: Compared to subjects without tax research decision aids, subjects provided justification requirement decision aids will place less importance on tax cases supporting their recommendations.

H2b: Compared to subjects without tax research decision aids, subjects provided factor evaluation checklist decision aids will place more importance on tax cases supporting their recommendations.

METHOD

Subjects

One hundred forty-two tax professionals from the then-Big 5 and two regional accounting firms located in two large Midwestern cities participated. Subjects ranged in tax work experience from one to 31 years, with an average of 7.2 years. Thirty-five percent of the respondents were female; 65 percent were male. Twenty-three percent of the subjects were staff level, 21 percent were seniors, 44 percent were managers, and 12 percent were partners. Fifty-one percent of the subjects had bachelor's degrees, 33 percent had graduate (master's) degrees, and 16 percent had J.D.s. Eighty-seven percent had passed the CPA exam. Big 5 subjects represented 81 percent of the sample, while regional firm subjects represented the remaining 19 percent of the sample.

Decision Aid Design

The design of the justification requirement decision aid is as follows. (1) A "balanced approach" is used, requiring users to list evidence both supporting and not supporting their assessments of the tax issue (Agoglia et al. 2003). (2) A blank text area of 1.5 letter-size pages is provided for subjects to write their justification statements. (3) Subjects are to

include in their statements the decision weights assigned to the balanced evidence and used to arrive at their assessments, and to explain why they assigned these weights. We expect design feature (1) to prompt subjects toward an objective appreciation of the evidence in sum because it requires including evidence contrary to their assessments. Feature (3) should guide subjects toward an integrated understanding of the tax issue because subjects have to consider the decision weights in relation to each other and estimate the amount of support or nonsupport provided by each piece of evidence. The "unrestrictiveness" of feature (2) allows subjects considerable freedom over statement content and form, which we expect to increase user interaction and encourage writing an integrated, noncomponent statement (Silver 1990). Using these features, the justification requirement used in this study is designed to enhance its ability to foster reflection on the task problem, reducing systematic judgment errors.

The design of the factor evaluation checklist is as follows. (1) Using a "component approach" (i.e., one that requires subjects to delineate and separately discuss the different components of the task), subjects are to evaluate each key factor (i.e., component) individually in separate sections (Agoglia et al. 2003). (2) Subjects receive no instructions to relate factors to each other. (3) The factors to be discussed are preselected for subjects. (4) Subjects are to indicate on bar scales and explain in a restricted text area (four lines) the degree to which each factor supports the alternative tax positions. We expect design features (1), (2), and (4) to impair subjects' achieving an integrated understanding of the tax issue. Subjects are prompted to assess each factor separately (and not relative to other factors), not required to write an integrated statement, and provided text areas too small to attempt integrating the factors, even if desired. Features (3) and (4) are expected to reduce user interaction by not allowing choice and by limiting the content and form of subjects' responses. The factor evaluation checklist is designed to enhance its tendency to efficiently guide users through the decision-making process, reducing cognitive effort.

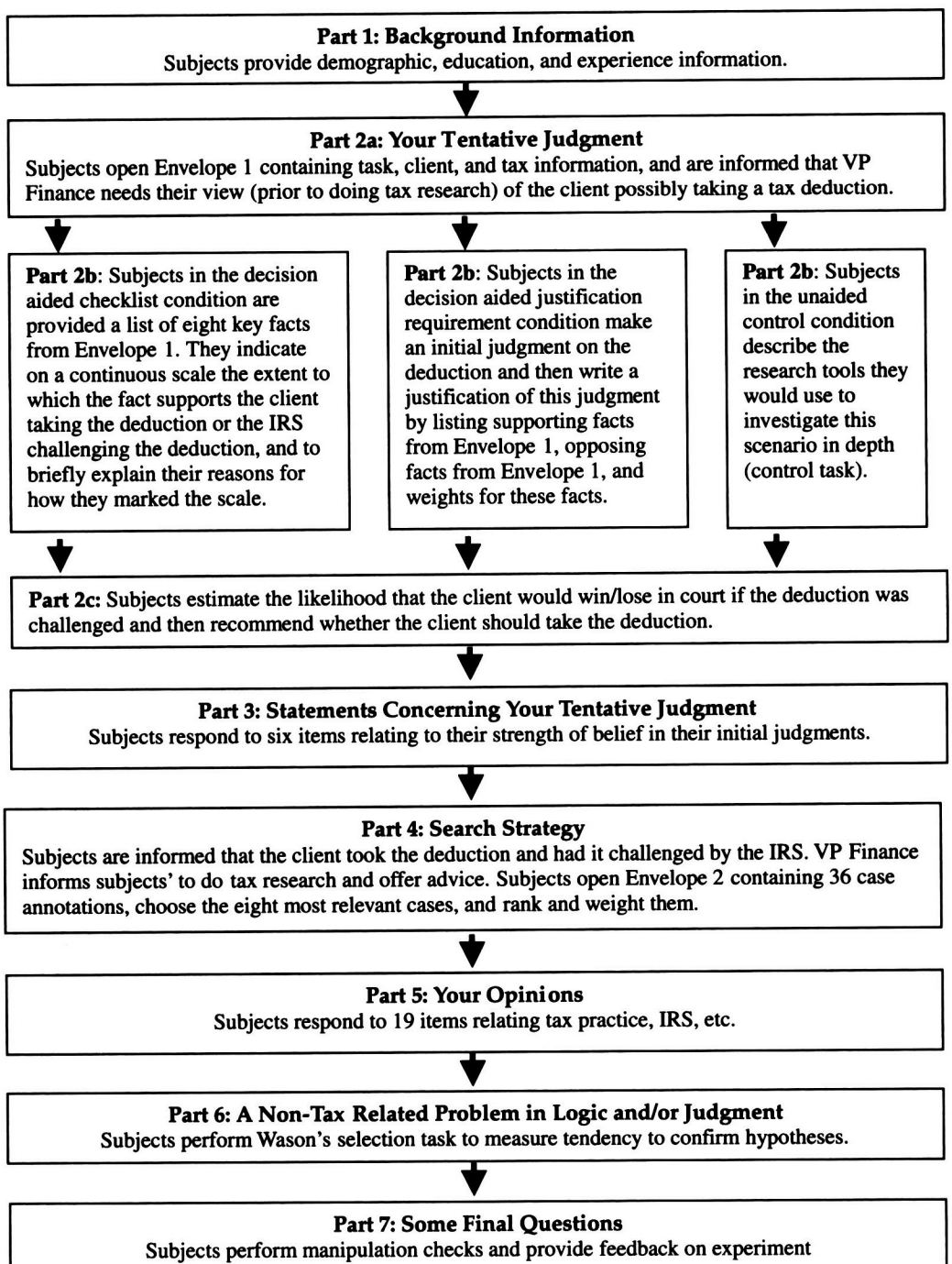
Experimental Design and Procedures

The task concerns corporate taxation and whether a pay bonus can be taken as a deduction. The task specifically relates to tax research, i.e., collecting and assessing tax evidence in relation to a tax issue. Our task materials were based on Johnson (1993), extensively modified and updated. Subjects performed the tax task by completing a multipart questionnaire tailored to their randomly assigned decision aid condition: justification requirement (n = 48), factor evaluation checklist (n = 49), and unaided (n = 45). Average time to complete the instrument was 75 minutes. To the extent possible within time and facility constraints, the task questionnaire mirrored a real-world tax research project involving planning and compliance, and recorded selected responses thereof. Big 5 tax practitioners extensively reviewed the study during pilot runs and found it realistic. The parts of the questionnaire are described in the following paragraphs and diagrammed in Figure 1.

In Part 1 (Background Information), subjects provided demographic, education, and tax experience information.[2] Part 2 (Your Tentative Judgment) consists of three phases, and required subjects in all conditions to express initial assessments of the task, i.e., (1) the likelihood or probability that the bonus, if paid, would be upheld in court as a deduction

[2] These data were collected to ensure randomization and to identify control variables. An analysis of the data using categorical models of subjects' gender, employment level, CPA status, and employer did not, however, indicate any significant differences in performance, supporting the effectiveness of the randomization process.

FIGURE 1
Diagram of the Experimental Instrument



to the corporation, and (2) a recommendation whether to pay the bonus and take the deduction.[3]

At the start of Part 2 (Part 2a in Fig. 1), subjects were instructed to read the contents of Envelope 1 which contained general information about the task, client, relevant tax codes, and "selected editorial comments from a major tax service." Subjects were not provided specific information from prior court cases, as in Part 4. Thus, the assessments in Part 2b below were made prior to conducting tax research.

Next in Part 2 (Part 2b in Fig. 1), subjects in the two aided conditions were provided decision aids, either the justification requirement or the factor evaluation checklist. Subjects using the justification requirement decision aid were asked whether it was "more likely" that taking the deduction would win or lose if challenged in court and to write out a balanced justification for this prediction. Specifically, they were asked to justify their prediction by listing from Envelope 1 facts that both supported and opposed the deduction, and to explain how they weighted the various facts in reaching their prediction. Subjects using the factor evaluation checklist decision aid were provided a list of eight key facts taken from Envelope 1 relating to the tax issue and asked to indicate their opinion as to the extent that each fact, considered independently, supported taking or not taking the deduction. Subjects in the unaided group were given a control task (indicate what research tools they would use if researching this situation for an actual client) in order to make both the activity and time required to complete the experiment comparable across all conditions.

In Part 3 (Statements Concerning Your Tentative Judgment), subjects expressed agreement/disagreement with six statements regarding their initial prediction of judicial success in Part 2. Key phrases in these statements were "sincere hope," "would be disappointed," "don't really care," "would please me," "feel committed," and "have no interest." These responses capture attitudes to the prediction made in Part 2b.[4]

In Part 4 (Search Strategy), subjects were informed that (a) the client decided to pay the bonus and take the deduction, and (b) the IRS challenged the deduction. As a result, the client's V.P. of Finance requested advice from the subjects on the situation as described in (a) and (b). Subjects were instructed to open and read Envelope 2, which contained 36 annotations (summaries) of court cases involving claiming bonuses as deductions. The 36 annotations, arranged in random order,[5] were based on nine different issues, only eight of which were relevant to the task case. Each case annotation reported one or two issues and indicated whether the court ruled in favor of the taxpayer taking the deduction. Of the 36 cases, subjects selected eight that they believed most warranted additional tax research in order to offer informed advice concerning the client's tax predicament. Subjects first listed the eight cases selected in any order and then ranked them in order of importance for tax research. Finally, subjects assigned "points" to the eight cases to indicate the decision weights they associated with the cases. The most important case was given a weight of 100 to serve as an anchor from which subjects could determine the relative values (0–100) of the remaining seven cases.

---

[3] Subjects were required to make an either/or choice concerning the deduction. While there may be more options than these two in practice, we limited our subjects in this manner to make any occurrences of confirmation bias measurable. Big 5 tax practitioners during pilot runs and 80 percent of the tax practitioner participants assessed the task as realistic.

[4] These variables proved to be insignificant in subsequent analysis and are retained in the remainder of the study only as covariates.

[5] Only one variation of the order of annotations was used, although the order within this variation was randomly determined. It is therefore possible that there may have been order effects not controlled for.

In Part 5 (Your Opinions), subjects responded to 19 statements relating to the IRS, tax code, tax practice, and court tax trials in general. None of the statements related specifically to the task or annotations in Part 4.[6] In Part 6 (A Non-Tax Related Problem in Logic and/or Judgment), subjects solved a Wason selection task (Eysenck and Keane 2000). This task is used to evaluate subjects' general tendency to confirm hypotheses.[7] Finally, in Part 7 (Some Final Questions), subjects answered six questions to verify that they were aware of certain facts in the task. All subjects passed these manipulation checks.

Subjects' Responses

Initial Prediction of Taxpayer Success. This variable is the subject's likelihood estimate of taxpayer judicial success, if the deduction was taken and then challenged in court. The estimates, made in Part 2, range from win as 100 percent to lose as 0 percent.

Number of Supporting Cases Selected. Of the eight cases selected for further tax research in Part 4, this variable is the number of those eight cases (0–8) supportive of the subject's recommendation made in Part 2 to take or not take the deduction.

Evaluation of Supporting Cases Selected. Based on the ratings of importance of the eight cases selected for further research in Part 4, this variable is the average (0–100) evaluated importance of selected cases that support the subject's recommended tax position from Part 2. Accordingly, of the eight cases, the number of them included in the calculation (0–8) is determined by the recommendation, e.g., for subjects recommending taking the deduction, of the eight selected cases, only the ratings of importance of those cases that had pro-taxpayer judgments were used to calculate the *Evaluation of Supporting Cases Selected* variable.[8]

RESULTS

Table 1 provides a descriptive summary of the main variables in this study. Although not hypothesized, we assumed that subjects would exhibit confirmation bias. Thus, when selecting in Part 4 the eight cases most relevant to conducting additional research on the tax issue, we expected subjects to select more cases supporting their recommended tax position (made in Part 2) than cases not supporting it. To test this expectation, we first classified case selections as either supporting or nonsupporting each subject's recommendation made in Part 2b. We used a single-sample binomial test, which combined the reciprocally related supporting and nonsupporting classifications for the two recommendation

---

[6] Similar to the background information collected in Part 1, we analyzed the data from Part 5 to identify potential control variables. Various analyses revealed no significant relationships of interest. Ten of the questions in Part 5 were based on Johnson (1993) and relate to client advocacy. We tested whether our subjects' degree of client advocacy influenced their initial predictions of taxpayer judicial success and found no significance. This is interesting because our experimental design resembles the design in Johnson's (1993) study, which found significant (although weak) advocacy effects. We also tested H1a/H1b and H2a/H2b with and without the advocacy variable as a covariate. The results were qualitatively the same. Because advocacy was never significant, we parsimoniously present our results without advocacy.

[7] Analysis of data from the Wason task revealed nothing of interest. For this lack of results, note Eysenck and Keane's (2000) discussion of the controversy as to whether this task measures a general cognitive bias toward confirming hypotheses or a more specialized bias affecting deductive reasoning.

[8] These ratings were standardized so that each subject's score could be reliably compared with the others' scores, regardless of the number of cases included in the calculation (Ghiselli et al. 1981; Johnson 1993). This controlled for subject effects, thereby allowing a person with a tendency to provide higher ratings in general to be comparable to a person with a tendency to provide lower ratings in general, and vice versa.

TABLE 1
Descriptive Statistics

	Overall (n = 142)	Unaided Condition[a] (n = 45)	Justification Condition (n = 48)	Checklist Condition (n = 49)
Panel A: Initial Prediction of Taxpayer Success[b]				
Range	9–95	9–95	10–90	20–95
Mean	63.5	64.7	65.1	60.8
Standard Deviation	(20.9)	(20.6)	(19.8)	(22.3)
Panel B: Number of Supporting Cases Selected[c]				
Range	0–8	0–8	0–8	0–8
Mean	4.8	5.2	4.1	5.0
Standard Deviation	(2.1)	(2.0)	(1.9)	(2.3)
Panel C: Evaluation of Supporting Cases Selected[d]				
Range	20–100	25–90	20–100	40–100
Mean	63.3	63.7	57.4	68.6
Standard Deviation	(16.8)	(16.2)	(17.6)	(16.4)
Panel D: *Strength of Belief* (Covariate)[e]				
Range	14–42	14–42	17–42	23–42
Mean	34.8	34.9	31.9	37.6
Standard Deviation	(6.4)	(6.6)	(7.3)	(5.2)

[a] *Decision Aid Conditions:* subjects were provided a justification requirement, a factor evaluation checklist, or no decision aid (as control benchmark).
[b] Initial Prediction of Taxpayer Success: subject's probability estimate of the client winning or losing in court.
[c] Number of Supporting Cases Selected: of the eight of 36 cases selected for further tax research, the number of cases supporting the subject's recommendation to the client to take or not take the deduction.
[d] Evaluation of Supporting Cases Selected: of the eight of 36 cases selected for further tax research, the evaluated importance of those cases supporting the subject's recommendation to the client to take or not take the deduction.
[e] *Strength of Belief*: covariate based on six questions concerning the subject's *strength of belief* to his/her recommendation to the client.

groups. The results of the test support the presence of confirmation bias ($z = 4.6$, $p < 0.0001$).[9]

We did separate analyses within the two types of recommendations (take or not take the deduction) because more cases with pro-taxpayer judgments were selected than cases with pro-IRS judgments. Cases resulting in pro-taxpayer judgments are defined as cases supportive for subjects recommending the deduction, whereas cases resulting in pro-IRS judgments are defined as supportive for subjects not recommending the deduction. This additional analysis was done to verify that the tendency to seek confirming evidence was not more dependent on one recommendation than the other. The results of this analysis were qualitatively the same as those in the above analyses. Subjects recommending taking

[9] Frequency and content analysis of the 36 cases suggests that no one case was favored over the others. The maximum any one case was selected was by 6.3 percent of the subjects, i.e., eight of the 142 participants.

the deduction preferentially selected pro-taxpayer cases, exhibiting positive search bias (i.e., disproportionately high selection of cases supporting the client's decision to take the deduction). However, those recommending against the deduction preferentially selected pro-IRS cases and thus did not exhibit positive search bias (but instead exhibited confirmation bias). That is, the occurrence of positive search strategies varied significantly by recommendation type, although confirmation bias occurred equally within both types of recommendation.

H1a/H1b and H2a/H2b predicted the effects of the two decision aid designs on the number and evaluated importance of the cases selected to support the tax practitioner's recommendation. Prior to testing these hypotheses separately, we analyzed the effect of decision aid condition on both the number and evaluated importance of selected cases. Table 2 presents the results of this analysis, indicating that decision aid condition was not significantly related to the number of selected cases, but is significantly related to the evaluated importance of selected cases. To properly interpret these results, it is necessary to test the two decision aids separately, as follows.

H1a/H1b pertain to the effects of the decision aids on the number of cases selected to support the tax researcher's recommendation. To test these hypotheses, the number of supporting cases was analyzed in an ANCOVA with *decision aid condition* as a categorical variable and *strength of belief* as a covariate.[10] *Strength of belief* is not significant in this or any subsequent test. H1a predicted that subjects with justification requirement decision aids will select fewer supporting cases than those without decision aids. Results support this prediction: the need for justification resulted in fewer supporting cases (and more nonsupporting cases) being selected ($F = 6.67$, $p < .01$). H1b predicted that subjects with the factor evaluation checklist decision aids will select more supporting cases than those without decision aids. This prediction is not supported ($F = 1.27$, $p = .29$).

H2a/H2b pertain to the effects of the decision aids on the evaluated importance of supporting evidence. To test these hypotheses, the evaluation of supporting cases was analyzed in an ANCOVA with *decision aid condition* as a categorical variable and *strength of belief* as a covariate, which is again not significant. H2a predicted that subjects with justification requirement decision aids will place less importance on supporting evidence than those without decision aids. Results support this prediction ($F = 6.29$, $p < .02$). H2b predicted that subjects with the factor evaluation checklist decision aids will place more importance on supporting evidence than those without decision aids. Results also support this prediction ($F = 5.32$, $p < .03$).

As additional testing of H1a/H1b and H2a/H2b, Scheffe comparisons were run from the two above ANCOVAs. Results from the Scheffe comparisons are shown in Table 2, Panel C.[11] For both the number of supporting cases selected and the evaluated importance of these cases, there were no significant differences between unaided condition and the factor evaluation checklist, whereas the justification requirement was significantly different from both the unaided condition and factor evaluation checklist. The Scheffe results and those in the previous two paragraphs agreed, except for the test of H2b (i.e., the Scheffe results indicated no difference between the checklist aid and the unaided condition, whereas

[10] *Strength of belief* is calculated from the six questions asked in Part 3 concerning subjects' affects toward their predictions of taxpayer success. Each item was scored on a seven-point scale and aggregated to form the *strength of belief* variable, resulting in a range of 6 (lowest) to 42 (highest). Cronbach's alpha for this set of questions was 0.89.

[11] We also ran Bonferroni comparisons and obtained the same results qualitatively as those reported in Table 2, Panel C.

TABLE 2
Effects of Decision Aid on Information Search

	Df	MS	F	p
Panel A: ANCOVA Results for Number of Supporting Cases Selected[a]				
Strength of Belief[b]	1	2.19	0.96	0.37
Decision Aid Condition[c]	2	1.41	1.34	0.25
Error	137			
Panel B: ANCOVA Results for Evaluation of Supporting Cases Selected[d]				
Strength of Belief	1	0.00076214	1.45	0.23
Decision Aid Condition	2	0.00403659	6.14	0.02
Error	137			

Panel C: Scheffe Comparisons from the above ANCOVAs

	Number of Supporting Cases Selected		Evaluation of Supporting Cases Selected	
Decision Aid Condition	**Mean**	**Grouping***	**Mean**	**Grouping***
Unaided (Control)	5.2	A	63.7	C
Factor Evaluation Checklist	5.0	A	68.6	C
Justification Requirement	4.1	B	57.4	D

* Means with same letter are not significantly different ($p \geq .05$); those with different letters are significantly different ($p < .05$). The two Scheffe comparisons are from separate ANCOVAs.
In the ANCOVA models, the *decision aid conditions* constitute the categorical variable, and *strength of belief* is a covariate. For cell means of variables, see Table 1.
[a] Number of Supporting Cases Selected: of the eight of 36 cases selected for further tax research, the number of cases supporting the subject's recommendation to the client to take or not take the deduction.
[b] *Strength of Belief*: covariate based on six questions concerning the subject's *strength of belief* to his/her recommendation to the client.
[c] *Decision Aid Condition*: subjects were provided a justification requirement, a factor evaluation checklist, or no decision aid (as control benchmark).
[d] Evaluation of Supporting Cases Selected: of the eight of 36 cases selected for further tax research, the evaluated importance of those cases supporting the subject's recommendation to the client to take or not take the deduction.

the initial test H2b did indicate a significant difference between these two conditions). Thus, in aggregate, these various tests suggest that the justification requirement aid had a significant effect in reducing confirmation bias compared to the other two conditions, whereas the factor evaluation checklist either had no debiasing effect or increased confirmation bias compared to the other two conditions.

SUMMARY AND CONCLUSIONS

We investigated the effects of two decision aid designs on confirmation bias in tax professionals deciding on information search strategies during tax research. Results indicate significant confirmation bias affecting the tax research. Interestingly, the confirmation bias was present during a compliance stage of the task, when one might expect confirmation bias to give way to positive search strategies supporting the client's final tax position. The bias was significant among participants regardless of whether they had previously recommended adopting the tax position under consideration.

H1a and H2a predict that a justification requirement decision aid would reduce confirmation bias in tax research. Both hypotheses are supported, indicating that this type of decision aid can reduce confirmation bias both in terms of the quantity of biased information selected and the importance attached to such information. Requiring tax practitioners to write balanced justifications of their recommendations can result in the selection of unbiased evidence, a better understanding of the tax issue, and a reduction of confirmation bias. Even in terms of supporting a client's position in court (i.e., client advocacy), it is important to examine prior cases in which the court found against positions similar to the client's in order to be ready to address these arguments should they be used by the IRS in the client's case.

H1b and H2b predict that a factor evaluation checklist decision aid would exacerbate confirmation bias in tax research. Results partially support these predictions: use of the checklist did not affect the quantity of biased information selected (contrary to H1b), but did significantly increase the importance attached to such evidence (supporting H2b). However, the results of H1b do not indicate that the factor evaluation check had a positive effect on decision-making. Similarly, Scheffe analysis indicates that decision-making with the factor evaluation checklist was no better than unaided decision-making and was significantly worse than decision-making with the justification requirement. Thus, providing tax practitioners with component-type decision aids that guide them through a mechanical step-by-step evaluation of the evidence can inculcate insular evaluations of the evidence, a fragmented view of the tax issue, and exacerbate or, at least, not counteract the biased selection of evidence.

The effects of the two decision aids on information search and confirmation bias are consistent with our predictions for the various design features examined: balanced/unbalanced, integrated/component, restricted/nonrestricted, and high/low interaction. The justification requirement used nonrestricted features that promoted user interaction, the formulation of an objective and integrated statement, and the collection of balanced evidence. The factor evaluation checklist used restricted and component-type features that impaired the assessment of the tax issue, user interaction, and the collection of balanced evidence. Using the above design framework to analyze existing tax decision aids and design new ones should facilitate predicting in a systematic manner the effects of tax decision aid features on tax judgment.

Nevertheless, our design of the two decision aids creates a limitation to this study in that the decision aids could have been designed differently. The factor evaluation checklist decision aid might have been designed to incorporate format features that would have allowed its use to improve judgment. The justification requirement decision aid might have been designed to impair judgment. However, we designed the two decision aids to reflect their typical characteristics and believe that our results indicate how these two types of decision aids usually affect tax information search. Future research might look at additional design features that would enhance the strengths and counteract the weaknesses of these two types of tax decision aids.

Another limitation of our study is that the task scenario is a mix of tax planning and tax compliance. Tax researchers might be expected to be more attached to their own ideas than they are to transactions already conducted by someone else. We did not separately measure confirmation bias in the earlier planning stage, but only in the compliance stage, because we believed it was most critical to investigate the persistence of an irrational confirmation bias after the client had adopted a tax position. Future research could examine confirmation bias during planning, as well as the effects of decision aids in this scenario.

Tax decision aids currently exist in a wide range of technologies (Masselli et al. 2002; Noga and Arnold 2002). Decision aid technologies include those as simple as instructions concerning easy-to-use formulas delivered through paper-based media (Arkes et al. 1986) and as complex as data-mining tools for web-based data warehouses (e.g., Vertex 2005), plus numerous technologies between these extremes (e.g., spreadsheets and hardcopy tax services). We investigated two tax decision aids that can be implemented over a substantial portion of the existing range of decision aid technologies. Although the two decision aids were provided to subjects in paper versions, the results concerning the justification requirement and factor evaluation checklist decision aids illustrate decision aid design principles that can be applied to all tax research methods, whether manual/paper-based, computerized/digitalized, or hybrid.

REFERENCES

Agoglia, C., T. Kida, and D. Hanno. 2003. The effects of alternative justification memos on the judgments of audit reviewees and reviewers. *Journal of Accounting Research* 41 (1): 33–46.

Anderson, C., M. Lepper, and L. Ross. 1980. Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology* 39 (December): 1037–1049.

Arkes, H., R. Dawes, and C. Christensen. 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. Sutton. 1998. The theory of technology dominance: Understanding the impact of intelligent decision aids on decision-makers' judgments. *Advances in Accounting Behavioral Research* 1: 175–194.

Ashton, R. 1992. Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes* 52: 292–306.

Barrick, J., C. Cloyd, and B. Spilker. 2004. The influence of biased tax research memoranda on supervisors' initial judgments in the review process. *The Journal of the American Taxation Association* 26 (1): 1–19.

Beck, P., J. Davis, and W. Jung. 1991. Tax advice and reporting under uncertainty: Theory and experimental evidence. *Contemporary Accounting Research* 13: 49–80.

Benbasat, I., and B. R. Nault. 1990. An evaluation of empirical research in managerial support systems. *Decision Support Systems* 6: 203–226.

Bonner, S. E., R. Libby, and M. W. Nelson. 1996. Using decision aids to improve auditors' conditional probability judgments. *The Accounting Review* 71 (2): 221–240.

Cloyd, C., and B. Spilker. 1999. The influence of client preferences on tax professionals' search for judicial precedents, subsequent judgments and recommendations. *The Accounting Review* 74 (3): 299–322.

———, and ———. 2000. Confirmation bias in tax information search: A comparison of law students and accounting students. *The Journal of the American Taxation Association* 22 (2): 60–71.

Cuccia, A., and G. McGill. 2000. The role of decision strategies in understanding professionals' susceptibility to judgment biases. *Journal of Accounting Research* 38 (2): 419–435.

Eysenck, M. W., and M. T. Keane. 2000. *Cognitive Psychology: A Student's Handbook*. Fourth edition. Hove, U.K.: Psychology Press.

Ghiselli, E. E., J. P. Campbell, and S. Zedeck. 1981. *Measurement Theory for the Behavioral Sciences*. New York, NY: W. H. Freeman and Company.

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

Hatfield, R. 2000. The effect of accountability on the evaluation of evidence: A tax setting. *Advances in Taxation* 12: 105–125.

———. 2001. The effect of staff accountant objectivity in the review and decision process: A tax setting. *The Journal of the American Taxation Association* 23 (1): 61–74.

Heider, F. 1958. *The Psychology of Interpersonal Relations*. New York, NY: Wiley.

Johnson, L. 1993. An empirical investigation of the effects of advocacy on preparer's evaluation of judicial evidence. *The Journal of the American Taxation Association* 15 (1): 1–22.

Johnson, V. E., and S. E. Kaplan. 1991. Experimental evidence on the effects of accountability on auditor judgments. *Auditing: A Journal of Practice & Theory* (Supplement): 96–107.

Jones, D. R., D. Brown, and P. R. Wheeler. 2001. The effect of justification type on agreement with a decision aid and judgment performance. *Advances in Accounting Behavioral Research* 4: 187–206.

Kachelmeier, S. J., 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.

Kahle, J., and R. White. 2004. Tax professional decision biases: The effects of initial beliefs and client preferences. *The Journal of the American Taxation Association* 26 (Supplement): 1–29.

Kahneman, D., P. Slovic, and A. Tversky. 1982. Shortcomings in the attribution process. *Judgment Under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.

Kennedy, J. 1993. Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research* 31 (Autumn): 231–245.

Magro, A. 1999. Contextual features of tax decision-making settings. *The Journal of the American Taxation Association* 21 (Supplement): 63–73.

Marshall, D., K. Misiewicz, and J. Parker. 1992. *Programmed Guide to Tax Research.* Fifth edition. Cincinnati, OH: South-Western Publishing Co.

Masselli, J., R. Ricketts, V. Arnold, and S. Sutton. 2002. The impact of embedded intelligent agents on tax-reporting decisions. *The Journal of the American Taxation Association* 24 (2): 60–78.

Mynatt, C., Doherty, M., and Tweney, R. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology* 29: 85–95.

———, ———, and ———. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology* 30: 395–406.

Noga, T., and V. Arnold. 2002. Do tax decision support systems affect the accuracy of tax compliance decisions? *International Journal of Accounting Information Systems* 3: 125–144.

Peecher, M. 1996. The influence of auditors' justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research* 34 (1): 125–140.

Plous, S. 1993. *The Psychology of Judgment and Decision-Making*. New York, NY: McGraw-Hill, Inc.

Rose, J. 2002. Behavioral decision aid research: Decision aid use and effects. In *Research Accounting as an Information Systems Discipline*, edited by Arnold and Sutton, Chapter 6. Sarasota, FL: American Accounting Association.

Rose, J. M., and C. J. Wolfe. 2000. The effects of system design alternatives on the acquisition of tax knowledge from a computerized tax decision aid. *Accounting, Organization and Society* 25: 285–306.

Ross, L., M. Lepper, F. Strack, and J. Steinmetz, J. 1977. Social explanation and social expectation: The effect of real and hypothetical explanations upon subjective likelihood. *Journal of Personality and Social Psychology* 35: 817–829.

Silver, M. S. 1990. Decision support systems: Directed and nondirected change. *Information Systems Research* 1: 47–70.

Tetlock, P. E., and J. I. Kim. 1987. Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology* 52: 700–709.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (September): 1124–1131.

Wheeler, P., and D. Jones. 2003. The effects of exclusive user choice of decision aid features on decision-making. *Journal of Information Systems* 17 (1): 63–83.

Vertex, Inc. 2005. Available at: http://www.vertexinc.com.