PROMOTING FLEXIBLE IDEA GENERATION IN AUDIT PLANNING: THE EFFECTS OF COUNTERFACTUAL MINDSET AND EXAMPLE PROVISION

BY

ELIZABETH ALTIERO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Accountancy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Mark E. Peecher, Chair
Associate Professor Clara Xiaoling Chen
Professor Ken Trotman, University of New South Wales
Assistant Professor Ravi S. Gajendran

Abstract

The creation of the Public Company Oversight Board and the release of their reports on audit deficiencies by public accounting firms have increased the dialog about what constitutes sufficient audit evidence. Since the determination of what is sufficient depends largely on professional judgment there is no bright line rule that auditors can apply to know they have completed a satisfactory audit. Using theory from the organizational behavior literature I examine several information gathering stopping rules that may be descriptive of how auditors make evidence sufficiency judgments (Browne and Pitts 2004). I perform interviews with practicing audit professionals to investigate the difficulties auditors face when determining the nature and extent of audit evidence that comprises a sufficient base on which to conclude. Results indicate that the audit program represents a *mental list* of the audit firm's collective knowledge against which individuals check off required audit evidence attributes. Through applying Bonner's (2007) Judgment and Decision Making in Accounting Framework, I identify significant differences in task, environment, and person variables across real world scenarios where the auditors found it was either very *easy* or very *difficult* to know when sufficient audit evidence had been collected prior to forming a conclusion.

In a second study, I run an experiment that manipulates task, environment and person (cognitive) variables identified in Study 1 to investigate how flexible idea generation can be improved during audit planning. In the face of regulatory and public scrutiny, audit professionals and researchers alike tend to focus on improving auditor decision making by reducing judgment errors. Doing so can lead to increased rigidity and stagnation of the audit program mental list, rather than incorporating changes that reflect new additions to collective firm knowledge. An alternative to improving performance through focusing explicitly on error reduction is to

encourage more insightful application of professional judgment (Klein 2013). Promoting flexible idea generation during the planning stage of the audit is also important because it can provide the unpredictability of procedures desired by regulators (AICPA 2006; PCAOB 2007; IFAC 2009) and can discover audit process improvements for future audits. Through application of the Search for Ideas in Associative Memory model (Nijstad and Stroebe 2006), I predict and find that priming a counterfactual mindset improves audit evidence idea generation when auditors perform familiar audit tasks by helping them overcome subconscious barriers created through repetitive use of standardized audit procedures. I also predict and find that provision of a suggested procedure improves audit evidence idea generation, but that priming a counterfactual mindset negates this benefit for auditors completing unfamiliar audit tasks. This study contributes to literature and practice by introducing flexible idea generation as an audit judgment performance measure for audit planning tasks beyond traditionally accepted fraud brainstorming procedures. Supplemental analyses also provide preliminary evidence that promoting flexible idea generation can improve audit efficiency for familiar tasks, which is indicative of flexible idea generation leading to audit process improvements.

Acknowledgements

I am grateful to my family and friends who have supported me throughout my time as a doctoral student. I appreciate all of the efforts of my dissertation committee, Ken Trotman, Clara Chen, Ravi Gajendran, and particularly my committee chair, Mark Peecher, for their comments and guidance throughout this process. Thank you to Will Ciconte and Brad Lang for their research assistance. This paper has also been improved by the thoughtful comments of Tony Bucaro, Andy Bauer, Tim Bauer, Cassandra Estep, Brooke Elliott, Stephanie Grant, Emily Griffith, Gary Hecht, Jeremy Lill, Aaron Zimbelman, and doctoral seminar and Accountancy Forum participants at the University of Illinois at Urbana-Champaign. Finally, I would like to thank the Accounting Doctoral Scholars program, the Richard D. and Anne Marie Irwin Foundation and the Victor L. Bernard Memorial Fund for the financial assistance they have provided.

Table of Contents

I. INTRODUCTION 1
II. STUDY 1: EXPLORATORY INTERVIEWS 4
III. STUDY 2: THEORY AND HYPOTHESIS DEVELOPMENT 27
IV. STUDY 2: EXPERIMENTAL METHOD 43
V. STUDY 2: RESULTS 51
VI. CONCLUSIONS 59
REFERENCES 64
FIGURES 71
TABLES 81
APPENDIX A: LIST OF INTERVIEW QUESTIONS FROM STUDY 1 103
APPENDIX B: SURVEY MATERIALS FROM STUDY 1 105
APPENDIX C: EXPERIMENTAL MATERIALS FROM STUDY 2 107

I. INTRODUCTION

The creation of the Public Company Oversight Board (PCAOB) and its inspection process for public accounting firms has shone a spotlight on the concept of audit evidence sufficiency. According to PCAOB reports, an audit deficiency occurs when, "at the time [the firm] issued its audit report, it had not obtained sufficient competent evidential matter to support its opinion on the issuer's financial statements" (PCAOB 2008, 7). In other words, audit deficiencies are the result of insufficient audit evidence collections. Auditing standards provide guidance as to what constitutes sufficient audit evidence. However, each auditee is unique and increased business complexity can result in even seemingly routine applications of sampling methodology requiring professional judgment (e.g., a multiple location entity with non-standard processes). In this thesis, I perform an exploratory interview study on evidence sufficiency judgments with professional auditor participants followed by an experimental study with interventions designed to address challenges observed in the interview study.

In Study 1, I apply theory related to information acquisition stopping rules in interviews with practicing auditors to examine what difficulties auditors encounter when attempting to determine whether they have obtained sufficient audit evidence. Fifteen auditors with ranks of Senior or In-Charge through Partner or Executive Director described, and completed brief surveys regarding, scenarios in which it was either very *easy* or very *difficult* for them to determine whether they had collected sufficient audit evidence. Browne and Pitts (2004) propose four different stopping rules, or information aggregation techniques, that are applicable to settings focused, "on the completeness or sufficiency of information obtained rather than on choosing between existing alternatives" (213). I find evidence that auditors prefer to use the Mental List stopping rule in which the auditor will assess sufficiency by comparing information

attained against a predetermined set of criteria. This is perhaps unsurprising given audit methodology structured around covering each assertion and audit programs that resemble giant checklists. Browne and Pitts (2004) also propose that more experienced individuals prefer the Mental List stopping rule, which is consistent with audit programs reflecting the collective experience (i.e., mental lists) of firm members. Further, I find that auditors had the most difficulty judging evidence sufficiency when the transaction being audited was unfamiliar to either the individual or the firm as a whole (i.e., no mental list was readily available).

Analysis of the survey responses, regarding attributes of the *easy* and *difficult* scenarios, through the lens of Bonner's (2008) Judgment and Decision Making in Accounting Framework informs my experimental design for Study 2. In this study, I investigate factors that can improve auditors' flexible idea generation regarding potential audit evidence that they might obtain. I examine flexible idea generation for two reasons: (1) brainstorming potential audit evidence provides the auditor with items from which they can build or revise a mental list; and (2) although the term evidence sufficiency naturally brings to mind concerns of quantity, the quality (i.e., type) of evidence collected can result in insufficiency as well. Using Bonner's classifications of task, environment, and person variables, I focus primarily on the task variable of routineness (conceptually, task familiarity), the environmental variable of working paper similarity (conceptually, provision of an example), and the person variable of experience (conceptually, relevant knowledge).

In Study 2, I investigate auditor flexible idea generation for two reasons: (1) in the *easy* scenarios of Study 1, flexible idea generation would encourage auditors to considering whether there was better audit evidence available than what their existing mental list what the audit program required them to obtain (i.e., firm's mental list); and (2) in the *difficult* scenarios

flexible idea generation would facilitate the application of auditors' knowledge and experience with unfamiliar and unusual tasks (i.e., those without mental lists). I integrate my observations from Study 1 with theory on idea generation to explore how a psychological intervention, priming a counterfactual mindset, can allow auditors to make better use of their experience and access more task relevant knowledge during audit planning. Additionally, I consider whether the audit environmental factors of example provision (e.g., availability of a prior year working paper) aids or hinders idea generation, both in isolation and in conjunction with the counterfactual mindset prime. Using theory from the Search for Ideas in Associative Memory (SIAM) model I predict and find that priming a counterfactual mindset provides greater benefits to idea generation for familiar tasks, regardless of example presence; alternatively, example presence provides the greatest benefits to idea generation in unfamiliar tasks, but only in the absence of a counterfactual mindset (Nijstad and Stroebe 2006).

I organize the remainder of this dissertation as follows. Chapter II describes the exploratory interview Study 1. Chapter III describes the theory and hypothesis development for Study 2. Chapters IV and V describe the experimental method and the results for Study 2, respectively. Section VI provides discussion and concluding comments.

II. STUDY 1: EXPLORATORY INTERVIEWS

> It's just judgment. …[I]t all comes down to judgment and your experience in working with the company, and it could be a different experience at a different company, and your risk tolerance, and your own personal experience working with a particular account, and…the trust you have in the competency of the client and the team. …[T]here's many, many factors.
>
> - Audit Partner 5

A practicing audit partner, when asked why two auditors might disagree on whether sufficient audit evidence had been collected for a given engagement, gave the above quote. Since the concept of evidence sufficiency can so easily vary from auditor to auditor, it is perhaps unsurprising that the PCAOB inspectors, with entirely different experiences, so often find and report audit deficiencies. While audit deficiencies related to quantity of evidence collected appear to be in decline, audit deficiencies related to the nature of evidence collected seem to persist (Church and Shefchik 2012). However, we know little about how auditors determine what evidence to obtain and how they apply professional judgment determine sufficiency.

I apply theory from the organizational behavior literature to examine how auditors determine when the have collected sufficient audit evidence. Browne and Pitts (2004) identify four information acquisition stopping rules that they theorize are most likely to be applicable to design problems (as opposed to choice problems). Information search for a design problem focuses on evidence sufficiency, whereas in choice problems the goal is to arrive at a correct answer. Therefore, information search strategies for design problems are more likely to be applicable to audit evidence collection.

The four stopping rules for design problems identified by Browne and Pitts are: magnitude threshold, mental list, difference threshold, and representational stability. The *magnitude threshold* rule posits that an individual maintains a mental calculation of the cumulative impact of all evidence gathered and will stop when the sum of the evidence surpasses

a predetermined amount. The *mental list* rule posits that a person has a criteria set in their mind against which they asses the sufficiency of information collected. The *difference threshold* rule posits that a person evaluates the incremental value of the most recent item obtained and stops searching for additional information when no longer learning anything new. The *representational stability* rule posits that a person continues to collect information as long as the new information adds something to their mental model of the problem.

I conducted interviews with practicing or former public accountants to gain an understanding of which rule or rules best represent auditors' evidence sufficiency judgments. The interviewees recounted experiences in which it was either *easy* or *difficult* for them to know when they had collected sufficient audit evidence. Throughout these conversations I noted a few overarching themes. In general, the Mental List stopping rule seemed to best represent the scenarios described. The audit program serves as something of a collective mental list built upon the expertise of many members of the firm. In some cases audit programs were described as being well defined and seemingly infallible.

> I would [compare] that to that technical work program that I was talking about. I want to see these attributes and once I see these attributes, I'm done. Not, "Okay, I've seen these attributes. Is there anything else that I necessarily need? "
>
> - Audit Partner 1

While other interviewees seemed to acknowledge that following the audit program blindly is not always adequate.

> [J]ust by following the audit programs, …they're rather detailed. They still do contain some level of subjectivity based on the risk of the client…but generally speaking, … making it through the audit program makes me comfortable.
>
> - Audit Senior 1

However, even when no audit plan existed, interviewees spoke of applying a basic mental list as a starting point.

> [W]ith any process…you're going to go through and say, "Alright, I've got the transaction. I've understood the transaction. I've tested the inputs and…tested the calculations."
>
> - Audit Manger 1

Overall, for the easy scenarios participants tended to describe tasks, such as cash reconciliations, that varied little over time and from client to client and therefore have well developed audit programs. For the difficult scenarios, however, participants tended to describe unusual transactions or events for which the audit program was not as well developed.

The *easy* and *difficult* scenarios differed in a number of other ways, in addition to the level of audit program development. In the *easy* scenarios the audit program was detailed, easy to implement and very similar either to procedures performed in prior year or on another auditee. In the *difficult* scenarios the transaction was nonstandard and sometimes unique, required significant amounts of professional judgment, and involved a high level of subject matter expertise. The detailed analysis of these differences informs my experimental design for Study 2.

Interview Method

Participants

Study 1 analyzes interviews conducted with one former and 14 practicing public accountants. All of the participants had experience in public accounting as financial statement auditors. I selected participants with varying ranks (average experience) such that I spoke with five Senior or In-charge auditors (henceforth, Senior) (3.4 years), five Manager or Senior Managers (henceforth, Manager) (8.4 years) and five Directors or Partners (henceforth, Partner) (22.4 years) (Table 1).[1] Across all ranks the participants possessed an average (standard deviation) of 11.4 (9.8) years experience. Of the 15 total participants 12 had experience at Big 4

[1] Participants with the rank of Partner reported a significantly greater years experience as an auditor than either the Senior ($F_{1,12}$= 29.79; $p < 0.001$, two-tailed) or Manager ($F_{1,12}$= 16.17; $p = .002$, two-tailed) ranks. There was not a significant difference in experience between the Senior and Manager ranks ($F_{1,12}$= 2.06; $p = .176$, two-tailed)

accounting firms while the remaining three participants gained their auditing experience at a regional midsize firm. The participants possessed a range of industry specializations including Financial Services (6), Insurance (3), Real Estate (3), Manufacturing (2), and Public Sector (1).

Interview Structure

The purpose of these interviews was to gain an understanding of the conditions under which it is difficult for financial statement auditors to know whether they have collected sufficient audit evidence upon which to base their judgments. These semi-structured interviews were administered either in person or by phone by the author and lasted 43.6 (7.9) minutes on average (standard deviation). All of the interviews were transcribed to facilitate analysis and these transcripts averaged (standard deviation) 6,983.0 (1,597.8) words.

Similar to previous accounting interview studies the primary interview questions required participants to recall previous audit experiences and relate the scenario, in detail, to the researcher (Gibbins and Newton 1994; Gibbins et al. 2001). Specifically, participants recalled two separate auditing experiences, one in which it was very *easy* to know when they had collected sufficient audit evidence and one in which this determination was very *difficult* (Appendix A, Primary Questions 1 and 2). The author randomized the interview materials prior to the commencement of any interviews regarding whether participants would describe the *difficult* or *easy* scenario first. Participants reported demographic information such as age and rank between scenarios as a distractor task. The purpose of the distractor task was to decrease participant anchoring on the first scenario described and to facilitate a greater degree of difference between the scenarios selected. Table 2 provides a brief description of the scenarios described by each participant.

At the conclusion of the conversation regarding each scenario, but before proceeding to the demographic or other interview questions, participants completed a brief survey instrument. The survey consisted of scale questions designed to capture differences between the two examples in a more quantitative way (see Appendix B for the survey questionnaire). At least two potential sources of bias exist for the scenarios participants selected. (1) The scenarios selected may be atypical since unusual situations are easier to recall. (2) The participants may select *very difficult* and *moderately difficult*, rather than *difficult* and *easy* scenarios as a method of ego bolstering (i.e., purporting that something that is difficult for most is easy for themself). I attempt to reduce the ill effects of these biases by focusing on differences in key characteristics between the scenarios within subjects, rather than on the level of each scenario between subjects. At the conclusion of both survey questionnaires participants had the opportunity to suggest any additional significant differences between the two scenarios that was not noted on the survey (Appendix A, Primary Question 3). The most common responses were (n): complexity of client, transaction, or accounting rule (3), estimate uncertainty (3), management competence (2), and audit or inherent risk (2).

The remainder of the interview questions solicited participant's opinions and observations about difficulty judging evidence sufficiency in general, rather than with regard to specific scenarios. Primary Questions four and six aimed to gather greater insight on applications of professional judgment and root causes of disagreements regarding audit evidence sufficiency (Appendix A). Primary Questions five and seven aimed to learn how auditors proceed when the audit program is incomplete or insufficiently specified (Appendix A). The author asked the Supplemental Questions when time remaining after the Primary Questions allowed. These questions aimed to explore some common audit research constructs such as reviewer vs. preparer

perspective and the use of red flags in identifying insufficient evidence. Finally, some of the participants provided real world audit tasks where they utilized each of Browne and Pitts (2004) four stopping rules (Appendix A).

Results and Analysis

To identify themes across the scenarios provided by the interviewees regarding what made the auditing task difficult, I perform a two-step analysis on the survey questionnaire responses. I first determine which questions elicited responses that varied significantly between easy and hard scenarios and then perform a factor analysis on the responses to the *difficult* scenario for the questions identified in Step 1. Using repeated measures ANOVA, I examine the effects of scenario difficulty and auditor rank, while controlling for order in which the interviewee described the scenarios (i.e., *easy* first or *difficult* first). I find a significant main effect of difficulty for all but three of the twelve questions. Accordingly, Question 7, the amount of evidence collected personally ($F_{1, 9} = 4.16$; $p = 0.069$, two-tailed), Question 8, the degree to which management offered evidence to the auditor ($F_{1, 9} = 0.06$; $p = 0.815$, two-tailed), and Question 12, the relationship between the auditor and management ($F_{1, 9} = 2.10$; $p = 0.181$, two-tailed), were excluded from further analysis. The second step of the analysis uses participant's responses to the remaining ten questions from the *difficult* questionnaire in a factor analysis that reveals four themes regarding what made the scenarios difficult to audit (Table 3).

Bonner's Framework for Judgment and Decision Making (JDM) Research in Accounting (2008) proposes three general classifications of factors that are likely to influence auditor JDM: task, environment, and person. These factors are descriptive of the conceptual factors identified in the survey questionnaire analysis. I discuss each factor individually in the following sections.

Task Variables

Task variables are factors that are characteristics of that can vary either between or within specific tasks auditors perform (Bonner 2008). In this research the tasks are the transactions that the auditors found either easy or difficult to know when they had sufficient audit evidence. I find that the survey questions load on two different task related factors. The four questions (factor loadings) that comprise the *task-internal* factor pertain to the amount of professional judgment required (0.98), how routine the transaction was for the client (0.92), the subjectivity of the audit evidence (0.67), and the average expertise of other audit team members (0.67) (Table 3). The first three items are all characteristics of the transaction being audited or of the related audit evidence. The final item, average expertise of other audit team members, is likely loads on the task variable because audit firms assign more experienced audit team members to non-routine, highly subjective transactions. The question (factor loading) that comprises the *task-external* factor pertains to task characteristics dictated by regulators through accounting guidance ambiguity (.98) (Table 3). I next consider each component of this factor in greater detail.

Professional Judgment

The auditor should use professional judgment in determining the nature, timing and extent of audit evidence that should be examined and also when evaluating the adequacy of that evidence prior to opining on the financial statements (AICPA 2006). Researchers have long understood the importance of better understanding auditors' use of professional judgment and the factors that influence its application (Joyce 1976; Gibbins 1984). More recently, the literature has focused on how the application of appropriate professional judgment becomes crucial to audit quality as the complexity of, and the uncertainty associated with, the transaction increases

(Peecher et al. 2013). Therefore, on average, the more professional judgment auditing a transaction requires, the more difficult it will be for the auditor to know when sufficient audit evidence has been collected.

Participants rated the amount of professional judgment required to audit the transactions described in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = No professional judgment used" and "10 = Extensive professional judgment used" while the midpoint was marked as "5 = Moderate professional judgment used" (Table 4). Participants rated the *difficult* scenarios as requiring significantly more professional judgment than the *easy* scenarios ($F_{1, 9} = 64.74$; $p < 0.001$, two-tailed). The difference in level of professional judgment required often seems to rely on the nature of the underlying evidence (i.e., subjective vs. objective) and the nature of the underlying accounting rule (i.e., historical cost vs. fair value estimation).

> Well, there's no judgment involved. [V]ersus an area of estimation…it's a very black and white transaction and…you're able to see the dollar amount involved. You're able to see the parties involved. You're able to see that everything matches up very clearly and…the dates match up…so you're able to conclude pretty easily.
>
> - Audit Partner 3

> [T]he example that I think people…could differ on vastly would be…security valuations and particularly on…these illiquid securities, these private label mortgage backs. I'm sure there's been instances where a client has supplied, "here's our information" and I'm sure there's been instances when some auditors have accepted that and others have asked to go further. …I think some…would live with management's judgments and estimates and some would feel like…they needed some outside validation.
>
> - Audit Partner 4

I find the effect of professional judgment consistently across auditor ranks. Senior ($F_{1, 3} = 45.07$; $p = 0.059$, two-tailed), Manager ($F_{1, 3} = 110.04$; $p = 0.002$, two-tailed), and Partner ($F_{1, 3} = 19.87$; $p = 0.021$, two-tailed) rank auditors all indicated at least a moderately significant difference of

professional judgment between the *easy* and *difficult* scenarios. Overall, the amount of professional judgment required was a major difference between scenarios when it was *easy* versus *difficult* to know whether sufficient audit evidence had been collected.

Routineness

Routineness in an audit context refers to the frequency with which a company engages in a particular type or size of accounting transaction (AICPA 2010). Routine processes tend to be central to the company's core operations including the primary revenue generating activities and administrative functions. These areas, because of their importance to the company and the volume of transactions processed are likely to have well established, and often automated, processes (AICPA 2010). Since routine processes are typically less complex than non-routine processes the auditor likely has established audit procedures that they perform year after year with little variation. Additionally, universally routine processes, such as payroll, tend to be fairly standardized across companies such that even new auditors to a client likely have experience auditing these transactions. Therefore, on average, the more routine a process is the easier it will likely be for the auditor to know when sufficient audit evidence has been collected.

Participants rated the routineness of the transactions being audited for both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = Routine" and "10 = Non-routine" while the midpoint was marked as "5 = Neither routine nor non-routine" (Table 5). Participants rated the *difficult* scenarios as significantly less routine than the *easy* scenarios ($F_{1,\ 9} = 11.56$; $p < 0.008$, two-tailed). One participant noted that non-routine transactions can create difficulties for auditors due to challenges obtaining all of the pertinent information so that the audit team can design an appropriate response.

> [T]he client explained that they were shipping product to a place that basically was still being built so it wasn't quite ready to receive the product. So they were

> going to ship it to another location, a third party warehouse, and there's specific revenue recognition guidance on when you ship to somewhere other than the ultimate destination in order to conclude that you have revenue. …[W]e had the company document specific facts as we understood them and then we could reach a conclusion. And so, in that…memo there were certain things…that they either hadn't conveyed to us, or conveyed slightly differently verbally. …[S]o we said, 'Can we see additional documentation?' and…we ultimately reached a different conclusion than we thought we would reach from the verbal conversation."
>
> - Audit Manager 3

Interestingly, I did not find this effect consistently across auditor ranks. Senior and Manager rank auditors both indicated that the *difficult* scenario was, with at least marginal significance, more routine ($F_{1, 3} = 9.94$; $p = 0.051$, two-tailed and $F_{1, 3} = 11.84$; $p = 0.041$, two-tailed, respectively). However, Partner ranked participants did not note such a difference ($F_{1, 3} = 0.03$; $p = 0.878$, two-tailed). For both the *easy* and *difficult* scenarios, the average partner response was below the midpoint of the scale, indicating that both scenarios described were routine in nature.

Although I do not make directional predictions ex ante, the findings at the Partner rank do appear counterintuitive. Had I made predictions *a priori* I would have expected the partners to would select extremely routine situations for the *easy* case and extremely non-routine situations for the *difficult* case. Alternatively, one might expect partners to select only more non-routine cases, as these would be places where their input was more often needed and would, therefore, be more salient in the partner's mind. In this case we might expect to observe a ceiling effect on this question. The results instead indicate what would be consistent with an interaction between experience and perceived routineness. The least experienced participants, seniors, gave responses indicating that they thought of routineness as whether they had received explicit training related to the transaction or not.

> We have a training program for how to audit cash and it's…the first thing you go to when you first started so it's pretty standard, routine and…you know people expect to see.
>
> - Audit Senior 3

The partners had significantly more years of audit experience than either of the other two rank groups (22.4 years vs. seniors' 3.4 years and managers' 8.4 years). It may be that perceptions of routineness are highly correlated with novelty or how often an auditor has encountered a transaction in the past. Given that the partners surveyed had an extra decade of experience over the other groups they are less likely to encounter a transaction they would deem truly novel and therefore may have an artificially high threshold for deeming a transaction non-routine. Consistent with this explanation is one partner's discussion of the subtle differences in fairly routine estimates at a financial services client.

> [W]ith the allowance for loan losses there, there's not a known, exact, "this is the right answer." …[O]ne could say there's not an exact right answer on the value of servicing assets, but I think you can look to more routine kinds of projections of values [or] discounted cash flows…you've seen many others that are in the same ballpark so you can feel comfortable with [the estimate]. [W]ith the allowance and the estimate of a loan loss it's, you know, just because someone else had a loan like that doesn't mean it's going to be the same result.
>
> - Audit Partner 4

Evidence Subjectivity

Audit evidence subjectivity varies widely from very objective items, such as third party bank statements, to very subjective items, such as management prepared estimates of warranty reserves. Accounting estimates, in general, are likely to be based on management's interpretation of the current environment and predictions about future events. This in turn results in audit evidence that possesses a high level of subjective elements, which are harder for the auditor to verify than objective facts (AICPA 1989). Therefore, on average, the more subjective the audit evidence is the more difficult it will likely be for the auditor to know when sufficient audit evidence has been collected.

Participants rated the subjectivity of the transactions being audited for both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = Entirely objective" and "10 = Entirely subjective" while the midpoint was marked as "5 = Equally subjective and objective" (Table 6). Participants rated the evidence in the *difficult* scenarios as significantly more subjective than in the *easy* scenarios ($F_{1, 9} = 9.34$; $p = 0.014$, two-tailed). In the difficult scenarios, participants spoke of estimates that were challenging to validate due to the subjectivity of the assumptions and the use of specialists as an attempt to evaluate reasonableness.

> Because valuations are very subjective…we were skeptical of the results. …Because it was a start up entity…you could have wildly different assumptions.
>
> - Audit Partner 5

> [T]here were two assumptions that [the valuation specialists] specifically helped with. …[T]he calculation is based on a couple…different methodologies and how you weight those methodologies is judgmental. Based on their experience and the number of valuations they've seen, they were able to provide some insight into…how they believe [the methodologies] should be weighted and the rational for that. …The second one…is the discount rate, which is built up over a combination of things and certain components - part of the discount rate is based on the risk free rate which is objective - other components are much more subjective. And so, based on their knowledge of the business compared to similar businesses that they've performed valuations on, they're able to see on the subjective component how much additional risk do our specific facts [add compared] to…companies that are similar in stature.
>
> - Audit Manager 3

Although only significant within the Senior rank ($F_{1, 3} = 10.40$; $p = 0.048$, two-tailed), I find results directionally consistent with this effect across all auditor ranks. Overall, the subjectivity of evidence was a significant difference between scenarios when it was *easy* versus *difficult* to know whether sufficient audit evidence had been collected.

Team Expertise

When auditing a transaction, auditors have the option to engage internal or external specialists that have greater subject matter expertise (PCAOB 2003). In the 1990's researchers noted a change in the organization of audit firms toward a model that emphasized industry specialized knowledge (Solomon et al. 1999). As auditee business models and the related transactions have continued to grow in complexity, understanding the use of both internal and external experts by audit teams has become critical (Griffith et al. 2014). Specialists are typically brought onto audits to review unusual, large, and/or high-risk transactions.

> [The firm's subject matter experts] have a lot of insight, and again we're talking to real estate specific people, so they have a lot more knowledge and wherewithal about what's going on in industry and what we see in industry practice. …So as opposed [to our]… smaller group here in Chicago, we were able to leverage off of what they've seen, talked about, heard from the SEC, and…leveraged their experience and their knowledge.
>
> - Audit Manager 4

Therefore, on average, audit firms are likely to assign audit teams with greater levels of expertise to audit transactions that are more difficult to determine audit evidence sufficiency. Some of the audit firms have cultivated reputations for being specialists in certain client industries in certain markets. One interviewee highlighted the need for a firm to ensure that they can provide appropriate team expertise.

> I wouldn't say it's common, but in first year audit clients we have had…disagreements with the prior audit firm…[on] the allowance, or some other significant estimate, and I don't know how to answer why their views were so different. I tend to think it is more related to their lack of expertise in an area that requires more specific knowledge than they possess.
>
> - Audit Partner 4

Participants rated the average expertise of other audit team members working on the transactions in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = No expertise" and "10 = Nationally recognized expert" while

the midpoint was marked as "5 = Moderate level of expertise" (Table 7). Participants rated the *difficult* scenarios as employing significantly more experienced audit teams than the *easy* scenarios ($F_{1, 9} = 9.40$; $p < 0.013$, two-tailed). Overall, the amount of professional judgment required was a major difference between scenarios when it was *easy* versus *difficult* to know whether sufficient audit evidence had been collected.

Additional analysis of the expertise results indicate that the effect is likely related to how audit firms assign professionals to individual transactions or tasks. The inclusion of routineness ratings as a covariate in the analysis negates the significant effect of expertise, but the expertise results are robust to the inclusion of professional judgment ratings as a covariate. Since estimates that require professional judgment to evaluate can be routine and quite common to certain clients and industries, these results indicate that the effect of expertise is likely due to audit teams selecting more experienced professionals for more unusual transactions or, in other words, the expertise of the team assigned is a reflection of task characteristics. One of the participants echoed this sentiment when discussing the difficult scenario.

> There were more people involved from the client side because of the technical nature of this. You needed certain people [on the audit team] that had various expertises.
>
> - Audit Manager 3

Accounting Guidance Ambiguity

Ambiguity in reference to accounting standards most typically refers to rules versus principles based accounting rules. Standard type has been found to influence auditors' perceptions of negotiation outcomes (Jamal and Tan 2010) and interact with auditor mindsets to influence financial statement quality (Ng and Tan 2003). Ambiguity can also be created through transformation of business models beyond what standard setters could have anticipated when originally drafting the accounting guidance.

> There there's an emerging revenue issue…[with] revenue earned from playing games on Facebook. …[T]he issue is, the total amount of revenue is known, but how you record that - whether you record what's called "gross" in the revenue line item or whether you record it "net" as an agent - is different. And so, because revenue from online social media…is a new concept, there's not authoritative guidance to directly address it. …[S]o how you reach the conclusion as to whether you meet the criteria to record it gross or record it net is based on applying your facts to [a set of] guidance that wasn't written with [social media games] in mind. …I can tell from publicly filed documents [of two companies with these transactions], neither one of which are my client, [that] the same set of facts are currently being treated differently.
>
> - Audit Manager 3

Increased ambiguity in a standard would likely make it more difficult to know what audit evidence would provide the greatest assurance. Therefore, when there is greater accounting guidance ambiguity the auditor will likely report it was more difficult to conclude on whether they have collected sufficient evidence.

Participants rated the amount of accounting guidance ambiguity that existed in both scenarios they provided (easy vs. hard to know when sufficient evidence had been collected) on an 11-point scale. The endpoints of the scale were "0 = Extremely unambiguous" and "10 = Extremely ambiguous" while the midpoint was marked as "Neither ambiguous nor unambiguous" (Table 8). Overall, participants did report that guidance ambiguity was higher in the scenarios where it was difficult to determine when they had collected sufficient audit evidence ($F_{1, 9} = 7.52$; $p = 0.023$, two-tailed). All ranks of auditors reported results that were directionally consistent with these combined findings. The Manager rank displayed the most significant difference indicating that these auditors likely encounter the greatest amount of variation in perceived guidance ambiguity on a daily basis. Manager rank auditors are often responsible for reviewing work done areas they are very familiar with and that have very little ambiguity (e.g., cash), but are also often the first to tackle accounting treatment questions on new, unfamiliar issues identified by the audit team or client management.

The analyses also indicated an order effect, related to which scenario the participant provided first (*easy* vs. *difficult*), that interacted with rank ($F_{1, 9} = 5.50$; $p = 0.028$, two-tailed). Further investigation of this interaction provided evidence that order effect was present in only the Senior rank. Senior participants rated scenarios to have significantly less accounting treatment ambiguity when describing the difficult scenario first ($F_{1, 3} = 159.16$; $p = 0.001$, two-tailed). The fact that the Senior rank shows an order effect where as the Manager and Partner ranks do not indicates that they are likely not well calibrated to how much ambiguity exists in the various accounting standards.

Environmental Variables

Environmental variables are factors that do not pertain to a specific individual or task, but instead relate to the conditions under which the individual must perform (Bonner 2008). The three questions (factor loadings) that comprise the *environment* factor pertain to management's outcome preference (0.87), the similarity of working papers to prior year (0.75), and budgetary pressures (0.75) (Table 3). While the first and last items are clearly environmental in nature and unrelated to the specific task, the second item is less obviously environmental in nature. Since I did not ask participants specifically why the working papers were less similar to prior year working papers in the *difficult* scenario I interpret this loading as environmental challenges being instrumental as to why the working papers were dissimilar. For example, anticipated regulatory scrutiny or internal quality review required reworking of what was done in prior years to meet reviewer expectations. I next consider each component of this factor in greater detail.

Treatment Conflict

When auditors and management disagree regarding the accounting treatment they must negotiate a mutually agreeable outcome or, if no agreement can be made, the auditor and auditee

will terminate their relationship. Accounting research on these negotiation processes focuses primarily on the interpersonal interactions between management and the auditor as well as the strategies they employ during those interactions (McCracken et al. 2008). The persuasiveness of the auditors' argument rests upon the evidence they are able to provide and since auditor evidence review is often constrained by cost and volume, it is important for auditors to identify the best evidence to support their position (Gibbons et al. 2001). It thus follows that in scenarios where management's accounting preference conflicts with the auditor it will be more difficult for the auditor to know when sufficient audit evidence has been collected.

Participants rated the amount of agreement with management regarding proper accounting for the transactions described in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = Management preference agreed with auditor" and "10 = Management preference conflicted with auditor" while the midpoint was marked as "5 = Management expressed no preference" (Table 9). Participants rated the *difficult* scenarios as more likely to have a conflict with management than the *easy* scenarios with marginal significance ($F_{1, 7} = 5.00$; $p = 0.061$, two-tailed). While this directional effect was observed across ranks, only the Partner rank reported a mean above the midpoint in the *difficult* scenario and the responses within the Partner rank appear to drive the overall results. This is likely due to the fact that lower ranks have less exposure to management pressure regarding accounting treatment conflicts. One partner participant described the difficulty experienced in resolving treatment conflict.

> It was not an easy answer. I consulted with our national office. I consulted with our SEC reviewing partner. I consulted with our risk management partner because it was such a significant…transaction…that I felt I needed to do that. …It was…hours and hours and hours of research and debate.
>
> - Audit Partner 5

Overall, these results suggest that accounting treatment disagreements with management vary between scenarios when it is *easy* versus *difficult* to know whether sufficient audit evidence had been collected.

Working Paper Similarity

Auditors are sometimes criticized for preferring to audit the same as last year, or SALY, through performing the same audit procedures year after year and using prior year working papers as a guide (Wright 1988). This technique is ineffective, and sometimes unavailable, in dynamic business environments. For auditees with evolving processes, obtaining the same audit evidence as prior year is likely ineffective, inefficient, or both. Auditors' general preference to audit SALY, despite potential detriments to audit quality, provides evidence that more stable business environments are likely less difficult in terms of determining evidence sufficiency. Therefore, working papers will likely be less similar to prior years in the *difficult* scenarios than in the *easy* scenarios.

Participants rated the similarity of the working papers to prior year's working papers for the transactions described in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = Identical to prior year" and "10 = No resemblance to prior year" while the midpoint was marked as "5 = Moderately similar to prior year" (Table 10). Participants rated the *difficult* scenarios as having significantly less similar working papers to prior year than the *easy* scenarios ($F_{1,\ 8}$ = 11.35; p = 0.010, two-tailed). This effect is directionally consistent across all ranks for auditors. Overall, working papers were less similar to prior year when it was *difficult* to know whether sufficient audit evidence had been collected.

Budget Pressure

Auditors are constrained by time budgets as to how much audit evidence they can collect and examine. Prior research indicates that greater amounts of time budget pressure result in decreases in the extent and depth of testing performed by auditors (Asare et al. 2000). When budget pressures are high overall the selection of the best possible audit evidence becomes more crucial. Additionally, decisions to continue to collect more evidence once already over budget must be justified to the audit team and management. It follows then that it would be more difficult to know whether sufficient evidence had been collected when auditors are faced with these pressures to justify their conclusions. Therefore, the audit engagement as a whole will likely be more over budget in the *difficult* scenarios than in the *easy* scenarios.

Participants rated the status of the engagement's budget during work on the transactions described in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = Extremely under budget" and "10 = Extremely over budget" while the midpoint was marked as "5 = On budget" (Table 11). Participants rated the *difficult* scenarios as being significantly more over budget than the *easy* scenarios ($F_{1, 9} = 21.42$; $p = 0.001$, two-tailed). This effect is directionally consistent across all ranks for auditors, but is driven by the Partner rank participants. Overall, audit engagements were more over budget when it was *difficult* to know whether sufficient audit evidence had been collected.

Participants across all ranks also identified budget pressure as a reason that an auditor might not collect sufficient evidence.

> I would say…budget constraints would be a big one. …Ultimately that's probably the top one.
>
> -Audit Senior 4

> I guess I would say, probably they didn't have enough time. Or they didn't request [documentation from the client] early enough. Or maybe it was an issue

> that came up last minute and there hadn't necessarily historically been issues on [that] client so maybe they were a little bit more trusting. More so probably just timing of it, whether it's budget constraints or just deadlines to wrap up the engagement, etc.
>
> - Audit Manager 1

> Some might say maybe they've already spent the budget, but I would choose to say they probably believe the client is trustworthy.
>
> - Audit Partner 4

Person Variables

Person variables are factors that relate to characteristics of the individual auditor as a decision maker (Bonner 2008). The question (factor loadings) that comprised the *personal* factor pertains to the amount of personal experience the auditor had with similar transactions (0.96) (Table 3). Experience is very closely related to topical knowledge, which varies across individual auditors and is specific to the task at hand. Tasks that an auditor has a great deal of knowledge about will, in general, be less difficult than tasks for which the auditor possesses little relevant knowledge.

Experience

The amount of experience auditors have with a type of transaction or an audit area will influence their ability to determine when they have collected sufficient audit evidence. Prior research has found that auditors with greater experience perform better on more complex tasks (Simnett 1996; Alissa et al. 2014). According to Browne and Pitts (2004) more experienced professionals can generate more effective mental lists. Auditors with greater experience will be able to better evaluate a transaction for similarities with work they have performed in the past and use evidence sufficiency criteria that have been successful in those scenarios.

> I think another thing would be…based on my own experience with whatever it is we're auditing. …[S]o someone says, "We can't get a system generated list from X application." [I think,] "No, my client over there does. So I know that you can do it!" So that would be another reason that I think I would say, "You know what,

> no." …[J]ust generally if it's something that we maybe haven't either asked for or we've struggled to get in the past I would push again and say, "Let's try to think of another way to get this information," as opposed to just [letting the audit client tell] us, "No, it's not going to work. It's not going to work. It's not going to work."
>
> - Audit Partner 1

> I think the experienced person is going to know, or have in their head, what he or she wants to find. [For example,] these three things and as long as these three things are there, or these three things make sense or, …are related to each other in the right way, then they'll be satisfied. Whereas the inexperienced auditor, …hopefully they know that they don't know…and that they should ask others that are more experienced as to what's sufficient.
>
> - Audit Partner 3

Therefore, auditors will likely report having less experience when describing *difficult* scenarios than when describing *easy* scenarios.

Participants rated the amount of experience that they personally had working on the type of transaction described in both scenarios they provided (*easy* and *difficult*) on an 11-point scale. The endpoints of the scale were "0 = No experience" and "10 = Extensive experience" while the midpoint was marked as "5 = Moderate experience" (Table 12). Not surprisingly, participant responses indicate that they had significantly less experience auditing the types of transactions described in the difficult scenarios ($F_{1, 9} = 12.76$; $p = 0.006$, two-tailed).

> [My manager is] more experienced in this so he would be able to not only more quickly identify the areas that we need to validate, but also…to tell on his own if certain areas…performed adequately or appear reasonable. Where [as] I might have to ask somebody…[and] dig into it a little more.
>
> - Audit Senior 4

These results are directionally consistent across all ranks with the Partner rank showing the smallest difference. For both scenarios, the Partner rank noted fairly high levels of experience, which indicates that this question may capture a ceiling effect of the extensive auditing experience of these participants. Overall, the results provide evidence that experience in auditing a certain type of transaction aids the auditor in making judgments about evidence sufficiency.

Conclusions

I find that of the four stopping rules identified by Browne and Pitts (2004), auditors most closely adhere to a *mental list* rule when judging evidence sufficiency. In practice, the audit program serves as a mental list representing the collective knowledge of the audit firm. When the audit program attributes were adequately specified, auditors described evidence sufficiency decisions as relatively *easy*. However, when audit program specifications were unclear, inadequate, or unavailable, auditors described evidence sufficiency judgments as relatively *difficult*. I then performed further analyses regarding the characteristics of scenarios that fell into each category (i.e., *easy* vs. *difficult*).

The results of the interview analyses indicate that all three types of variables in Bonner's (2008) Judgment and Decision Making in Accounting Framework differ across scenarios where it was *easy* versus *difficult* for auditors to know whether they had collected sufficient audit evidence. For *task* variables, auditors reported *difficult* scenarios required more professional judgment, related to transactions that were less routine, required the examination of more subjective evidence, utilized more experienced audit teams, and the accounting guidance was more ambiguous. For *environmental* variables, auditors reported *difficult* scenarios had more disagreement with management over the appropriate accounting treatment, less similarity in the final working papers compared to prior year, and greater time budget pressure. For *person* variables, auditors reported having less experience in the *difficult* scenarios.

I use the above observations to inform my design of Study 2. One of the more interesting observations from Study 1 was the interaction between participant rank and the perceived routineness of the *easy* vs. difficult transactions described. I posit that because of partners' vast experience, it is less likely partners will encounter transactions that are truly novel and this

decreases overall perceptions of routineness.[2] In other words, increased experience has made transactions appear *familiar* to partner ranked auditors than do to more junior ranked auditors. For the more junior auditors, *unfamiliar* transactions pose a greater challenge in determining how their auditing knowledge can be applied to collect sufficient, competent audit evidence. In Study 2, I apply the Search for Ideas in Associative Memory model to investigate how and to aid auditors in accessing a greater amount of relevant knowledge when auditing both familiar and unfamiliar transactions.

[2] It is important to note that truly unique transactions do occur, regardless of perceived routineness. For example, Lehman Brothers' Repo 105 transactions were unique and failure of the auditors to fully appreciate this likely played some part in Lehman's collapse.

III. STUDY 2: THEORY AND HYPOTHESIS DEVELOPMENT

Regulatory scrutiny in the wake of recent accounting scandal places increased pressure on auditing firms to emphasize error reduction and to, arguably, restrict individual auditors' professional judgment. Other professional organizations, such as the Central Intelligence Agency and the British Broadcasting Corporation, similarly chose to focus on error reduction in the wake of publicized failures, but have since observed instances where overall performance was impaired by not encouraging the flexible idea generation of employees (Klein 2013). The creation of the PCOAB and the release of their inspection findings has highlighted the importance of gaining a better understanding what it means to obtain sufficient audit evidence. Naturally, the word sufficient calls to mind questions regarding the quantity of evidence collected. Research shows that PCAOB deficiency findings that can be remedied by increasing the quantity of evidence obtained (i.e., audit sampling) appear to be in decline. However, judgment related deficiencies (i.e., fair value measurements) where the type of evidence can be more crucial to sufficiency than quantity appear to persist (Church and Shefchik 2012).

In Study 1 I find that auditors use the audit program as a mental list, which they use to assess the sufficiency of audit evidence collected. In the case of quantity deficiencies, standardization and error reduction tactics might increase sample size noted on the audit program or improve controls governing adherence to the audit program (e.g., restrict ability to change the audit program, clarify sampling policy, implement controls requiring use of sampling software, etc.). In the case of judgment related deficiencies the mental list (i.e., audit program) itself is likely deficient and it is therefore unclear how stricter adherence to the audit program will help auditors identify higher quality, and thus more sufficient, audit evidence. Additionally, audit firms that increase organizational control through standardization risk sacrificing the audit

teams' ability to provide required levels of unpredictability in audit procedures and their likelihood of identifying audit process improvements (AICPA 2006; PCAOB 2007; IFAC 2009; PCAOB 2010; Budescu et al. 2012). For example, standardized testing of information technology access controls requires auditors to select a sample of employees and manually check whether their system access is appropriate, but selecting this procedure because it is firm-recommended ignores technological advancements that now make it possible at some auditees to test 100% of employees in less time (Anonymous Audit Executive Director 2012).

Promoting flexible idea generation can also be beneficial to auditors facing relatively unfamiliar audit planning tasks. Auditors who are encouraged to exercise flexible professional judgment will be more successful at identifying commonalities between unfamiliar transactions and existing audit knowledge and responding accordingly when no standard audit program exists. For example, prior to the financial crisis there was no precedent for determining the appropriate audit evidence to collect when auditing a loss share agreement in which a bank purchased deposits and loans from the Federal Deposit Insurance Corporation (Anonymous Audit Senior 2012). Although audit firms and regulators are unlikely to dispute the need for encouraging innovative audits, particularly in light of a rapidly evolving business environment (e.g., increased use of big data), there is little incentive for any one auditor to consider alternatives to standard procedures on their audits even if explicitly told to do so (Low and Tan 2011; Arning 2013; Chipman 2013; Peecher et al. 2013).

Using the Search for Ideas in Associative Memory (SIAM) model from psychology to guide this investigation, I consider the potential benefits of priming a counterfactual mindset on audit evidence idea generation in two audit planning tasks where I also manipulate the presence or absence of an example standard audit procedure. Both accounting and psychology research

document that prompts encouraging a decision maker to engage in counterfactual *thinking* can improve judgment by increasing the amount and variety of information considered (Koonce 1992; Kray and Galinsky 2003; Galinsky and Kray 2004). Priming a counterfactual *mindset* will allow this increased information consideration to spill over onto subsequent tasks where subconscious barriers would otherwise hinder the effectiveness of counterfactual thinking (Galinsky and Moskowitz 2000). Since the SIAM model indicates that new ideas are generated based on activated domain specific knowledge, counterfactual primes can improve idea generation to the extent that the auditor possesses additional relevant, yet unconsidered, knowledge.

When auditing a familiar account or transaction, auditors are likely to access a very small subset of their auditing knowledge due to strong problem-knowledge links from repeatedly auditing similar transactions in a standardized way. These links can be difficult to break through effort alone due to their subconscious nature as evidenced by auditors' difficulty in a similar task, alternative hypothesis generation during analytical review procedures (Heiman 1990). I predict that priming a counterfactual mindset will aid in the activation of a wider range of task relevant knowledge and help auditors generate ideas in familiar audit tasks (Galinsky and Moskowitz 2000). Since standardized procedures are displayed prominently in the audit workplace through prior year work (e.g., Wright 1988; Low 2004) and regulator or firm guidance (e.g., Kennedy et al. 1997) it is important to investigate whether the predicted benefit of priming counterfactual thinking is robust to the presence of suggested example provision. I propose that it is not the suggested example itself, but rather it is the creation of a cognitive barrier from repeatedly using the same examples to audit familiar transactions in the same way that suppresses idea generation. Therefore, to the extent that the example is different from the

individual auditor's most common response, it will also be beneficial to idea generation through prompting a new line of knowledge search, but to a lesser extent than the wider knowledge search prompted by priming a counterfactual mindset (Nijstad et al. 2002; Nijstad and Strobe 2006).

When auditing an unfamiliar account or transaction, auditors are likely to encounter a different knowledge search challenge in attempting to identify what portions of their vast auditing knowledge are most relevant to the problem at hand. Here the issue for the auditor becomes narrowing the search of their knowledge to the most relevant audit related knowledge and therefore widening the search with a counterfactual mindset will likely be of limited value until the search becomes better defined. When examining an unfamiliar task that contains an example solution, a counterfactual mindset will increase auditors' focus on task-embedded cues and decrease the search for outside, related knowledge (Kray et al. 2006). The resulting ideas will be more similar to the example and less diverse overall. The provision of a standardized example in the absence of a counterfactual mindset, however, will be beneficial because it can act as a search cue that helps direct auditors toward relevant audit knowledge that would have otherwise remained untapped (Nijstad et al. 2002; Nijstad and Strobe 2006)

In a 2 x 2 x 2 experiment using audit professionals, I manipulate task (less familiar versus more familiar), counterfactual mindset prime (prime versus no prime) and example provision (present versus absent). Both experimental tasks asked participants to generate as many audit evidence ideas as they could for a hypothetical audit scenario. The more familiar case contained cues that clearly indicated a need to access revenue related audit knowledge. The less familiar case abstracted away from a traditional financial statement audit context and omitted the use of many common contextual cues that would indicate what type of audit knowledge would be

useful. I manipulate counterfactual mindset by having participants read a story that was designed to elicit, or not elicit, counterfactual thoughts prior to performing the audit task as in Galinsky and Moskowitz (2000) and Kray et al. (2006). For both tasks, the example, when provided, suggested that the auditors obtain data for use in a volume estimate analytical procedure. The primary dependent measure is a multidimensional evaluation of idea generation flexibility taken from the divergent thinking literature that evaluates the number, breadth, originality, and elaborateness of participants' audit evidence ideas (Guilford 1967; Karakelle 2009; Lewis and Lovatt 2013).

I find that, in both tasks, auditors not exposed to the example or counterfactual prime exercise less flexibility in idea generation than do auditors exposed to the manipulation better equipped to combat the knowledge search problem created by the task, despite all auditors receiving instructions to generate as many ideas as possible. In the more familiar task the results indicate that a counterfactual mindset helps auditors create audit evidence ideas that reflect a wider range of their knowledge, but this prompt is not beneficial in the less familiar task. In the less familiar case the provision of an example procedure does not inhibit the consideration of additional ideas and instead provides benefits of its own. However, these benefits disappear in auditors primed with a counterfactual mindset. Examining the overall pattern of results suggests that priming a counterfactual mindset is the most beneficial intervention to encourage flexible idea generation for a familiar transaction regardless of example presence, but that providing an example in isolation is the better approach for unfamiliar tasks. Supplemental analyses also suggest that the improved idea generation created by the interventions in the more familiar scenario allowed auditors to design more efficient audit programs (i.e., programs that required fewer estimated hours to complete).

This study contributes to the auditing literature by introducing flexibility in idea generation as a performance measure for audit planning and audit program design. Prior auditing research, like the audit firms and other professional organizations, generally has focused improving auditor judgment and decision-making by decreasing errors rather than by increasing insights (Klein 2013). One exception to this is fraud brainstorming where the idea of flexible idea generation is more accepted and encouraged by regulators, audit firms and researchers (Carpenter 2007; Hoffman and Zimbelman 2009; Trotman et al. 2009; Chen et al. 2013). Even within this work only Trotman et al. (2009) examines the originality of brainstormed frauds as benchmarked by other participants rather than ability to identify normatively correct responses. I expand and continue this stream of research by introducing a multidimensional idea generation measure and examining the benefits of flexible idea generation in additional areas of audit planning.

My study also complements recent accounting research that considers the interdependence of potential interventions (analogous to management controls) both with other interventions and with task or environmental characteristics (Grabner and Moers 2013). While prior auditing literature often considers decision aids in isolation (e.g., Kennedy 1995, Asare and Wright 2004; Carpenter 2007; notable exceptions include Hoffman and Zimbelman 2009 and Lynch et al. 2009), my study shows that the effectiveness of a frequently used decision aid, counterfactual thinking, depends on the use of other prompts such as example provision, as well as task characteristics.

Finally, this study also extends the psychology literature in the area of counterfactual mindset by identifying two previously unidentified factors, example provision and task familiarity, that jointly moderate the effect of a counterfactual mindset in idea generation tasks.

The presence of an interaction between counterfactual mindset and task familiarity indicates the need for additional idea generation research utilizing professional participants performing tasks in their areas of expertise and also highlights the need for additional research of idea generation in the audit domain.

Flexibility in Auditing and Accounting

Recent field research indicates that improving professional performance is likely a function of both reducing the occurrence of errors and the increasing the generation of insights (Klein 2013).[3] Much of the auditing literature to date, however, focuses on the former and ignores the latter (e.g., Libby and Libby 1989; Kachelmeier and Messier 1990; Bedard and Biggs 1991; Kennedy 1993; Messier et al. 2008). The fraud brainstorming area is one notable exception in which researchers, audit firms, and regulators seem to agree that promoting flexible idea generation is beneficial to audit quality.[4] The focus almost solely on error reduction outside of the fraud identification realm is puzzling given that highly mechanized audits face increased risk of audit failure due to predictability of audit procedures allowing management to anticipate what will be audited and how (AICPA 2006; PCAOB 2007; IFAC 2009; PCAOB 2010; Budescu et al. 2012).[5] One possible reason for fraud brainstorming's acceptance as a safe outlet of flexible idea generation is the mandate that it occur during the planning stage of the audit. At this time, the

[3] The constructs of insight and flexible idea generation are closely related to the construct of creativity. Researchers of creativity make the distinction between two different types of creativity that have varying degrees of applicability to an accounting and auditing setting. Big C refers to the more rare type of creativity that leads to major breakthroughs that are impactful to society. Little c refers to everyday problem solving, divergent thinking and the ability to adapt to change (Hennessey and Amabile 2010). The latter form of creativity should be particularly helpful to auditors attempting to adapt to the day-to-day challenges of ever changing clients and economic environments and is the primary focus of this research.

[4] During fraud brainstorming procedures, auditors generate a list of ideas regarding how management could commit fraud. The underlying premise if this exercise is that a longer and more diverse list will result in the design of procedures that are more responsive to the overall risk of fraud (Carpenter 2007; Hoffman and Zimbelman 2009; Trotman et al. 2009;Hammersley 2011; Chen et al. 2012; Herron 2012).

[5] In fact, even unpredictability that is, predictably, focused only in higher risk areas can unintentionally increase management's ability to predict where auditors will focus their efforts (Bowlin 2011), thus highlighting the potential value of more creative unpredictability in audit procedures across the entire engagement.

entire audit team comes together to vet ideas, which makes it a controlled environment to vet new audit procedures as well.

Auditors considering a wider range of potential audit evidence and procedures can provide similar benefits to fraud brainstorming through improved ability to respond to changes in macroeconomic and auditee specific risk. Research clearly documents a longstanding preference by auditors to audit the "same as last year" (Joyce and Biddle 1981; Wright 1988; Asare and Wright 2004; Brazel et al. 2004). Theory suggests this repetition of performing the same audit year after year can create a subconscious barrier that impairs auditors' ability to improve existing audit procedures and adapt to new developments at individual clients (e.g., process changes or new transaction types) or in the business environment as a whole (e.g., advent of big data or the housing market crash) (Cardoso and Badke 2011; Arning 2013; Chipman 2013).[6, 7] More recent literature further indicates that, particularly when facing unusual situations, auditors do not display high levels of divergent thinking without prompting (Plumlee et al. 2014). Therefore, the goal of promoting flexible idea generation during audit planning is to both increase consideration of alternative approaches to auditing old, familiar transactions and also to encourage better application of existing audit knowledge to new, unfamiliar transactions.

The Search for Ideas in Associative Memory (SIAM) model provides a framework that can inform predictions about what types of prompts or circumstances will stimulate flexible idea generation. This theory of memory recall is particularly relevant to professional settings as it

[6] Robert Arning, Vice Chair of Market Development at KPMG, LLC, gave a presentation in the Fall of 2013 where he stressed the importance his firm has placed on hiring creative people that will be able to keep up with increasing levels of innovation that the firm is seeing in their current and potential clients. Similarly, Stephen Chipman, Chief Executive Officer at Grant Thornton, LLP echoed Arning's sentiments at the 2013 American Accounting Association Annual Meeting when he spoke about his concern that auditors are not adapting quickly enough to changes in the business environment such as underutilizing Big Data.

[7] In addition to cognitive barriers caused by repetition, auditors face a number of potential obstacles to flexible judgment such as lack of individual incentives to design new audit approaches (Peecher et al. 2013), time pressure (Bonner 2008), and other subconscious cognitive barriers (e.g., production blocking in Carpenter 2007).

describes the process through which an individual interprets a problem and generates new ideas based on relevant topical knowledge, as opposed to models that ignore knowledge search and rely on other factors such as individual differences in creative ability (Raaijmakers and Shiffrin 1981; Nijstad et al. 2002; Nijstad and Stroebe 2006; Hennessey and Amabile 2010). Per the SIAM model an individual first examines a task for key information that can be used as a search cue to activate relevant knowledge and then use features of that activated knowledge to generate new ideas. In other words, auditors who are able to activate the greatest amount of *task relevant* knowledge will produce the greatest number and variety of audit evidence ideas for use in audit planning.

I describe below the challenges of idea generation in higher and lower familiarity tasks separately, beginning with the former.[8] When attempting to improve procedures the auditor will know what accounts the transaction should impact, what technical accounting rules are applicable and what audit procedures have been used in the past. These details of the task act as search cues that easily link the transaction description to relevant audit knowledge about how to approach the audit task. Difficulties in generating ideas, for these familiar tasks, stem from the existence of strong cue-knowledge-idea links from repeatedly auditing similar transactions in the same way (Figure 1, Panel A) (Nijstad and Stroebe 2006). Strong links of this type will suppress other knowledge and new ideas from coming to mind, thus prompts that increase the breadth of knowledge accessed by a search cue will likely increase the flexibility of idea generation.

[8] Familiarity is a construct that takes on many names in the accounting and psychology literature. For example, psychology has looked at situations in which experts face tasks in which the *content* is relevant to their area of expertise, but a lack of *contextual* information obscures this connection and inhibits memory search formation (Tversky 1977; Einhorn and Hogarth 1981). In accounting, the ease with which task cues can be linked to previously stored knowledge is called *input clarity.* In other words, a task with higher input clarity possesses a greater degree of similarity between the task cues and prior knowledge (Bonner 1994; Chen et al. 2014).

Alternatively, auditors can encounter a transaction that is unfamiliar to them and they must work hard to interpret the problem and form a search cue.[9] Ill-formed search cues, ones that do not properly interpret the problem and identify the type of knowledge needed, will likely call to mind knowledge that is irrelevant to the problem at hand and will block productive idea generation (Figure 1, Panel B) (Nijstad and Stroebe 2006). When irrelevant knowledge interferes with idea generation then prompts that are able to refine the search cue and facilitate accesses of a more appropriate subset of knowledge will likely increase the flexibility of idea generation. In sum, both familiar and unfamiliar tasks pose challenges to idea generation due to sub-optimal knowledge activation: (1) *more familiar* tasks activate knowledge that is too narrow and typically confined to previously learned routines and (2) *less familiar* tasks activate knowledge that is too broad and not confined to relevant topics.

Counterfactual Mindset Prime

Prompts to engage in counterfactual thinking have a well-documented effect of increasing the consideration of alternatives in both accounting and psychology research, making them a prime candidate for expanding activated knowledge. Counterfactual thinking is a form of mental simulation in which individuals use their imagination to consider what could have been in the past or what will be in the future (i.e., construct a mental scenario that is different from the current reality). The benefits of prompting counterfactual thought include increased

[9] In 15 semi-structured interviews, I conducted with practicing professionals participants spoke about situations they encountered that were either novel to them, their team, and/or the firm as a whole. One participant spoke about the difficulties in auditing and disclosing revenues associated with Facebook games because the best application of revenue recognition rules is unclear for such an intangible environment. Similarly, another interviewee spoke about a securities lending transaction that was quite simple to account for until the financial crisis resulted in the collateral, which was similar to a money market account where the net asset value is typically stable at one dollar, dropping below that valuation benchmark and requiring the audit client to add cash to the account. There was no standard audit plan available for use in this situation as the circumstances were not widely anticipated and continuing to audit according to the standard plan would have ignored the enhanced risk of the transaction under current market conditions. These examples provide evidence that auditors do encounter transactions that are low in familiarity during the course of their audits.

consideration of disconfirmatory evidence (Kray and Galinsky 2003), increased sharing of information in groups (Galinsky and Kray 2004), decreased effects of the curse of knowledge (i.e., the inability to disregard information already processed when it would be beneficial to do so) (Kennedy 1995) and increased skepticism of management's non-error explanations for analytical review procedures (Koonce 1992).[10] In all of these studies, prompts to actively think about a situation counterfactually made the participants less apt to follow their initial response and to instead incorporate a greater amount of information and knowledge into their judgments.

Auditors already attempting to generate potential audit evidence ideas will likely not be greatly influenced by a conscious prompt to think counterfactually and consider alternatives, as that is already their primary objective. Low and Tan (2011) find that instructions simply asking auditors to "think outside of the box" *do not* induce modifications to audit procedures in the absence of time pressure, which is typical of the environment during the planning stage of the audit.[11] The SIAM model indicates that this is likely due, in part, to common search cues and knowledge become linked in memory such that the repeated use of a search cue will subconsciously block the retrieval of different knowledge (Nijstad and Stroebe 2006). This theory is consistent with prior findings that auditors have difficultly generating multiple explanations during analytic procedures, even when instructed directly to do so (Heiman 1990). Psychology research has shown that an alternative form of the counterfactual prompt, priming a *counterfactual mindset*, is more successful at prompting the consideration of alternatives where

[10] Counterfactual thinking encompasses any consideration of an alternate reality, or thoughts of what might have been under different conditions. One technique use to elicit counterfactual thought is counterexplanation. Counterexplanation requires individuals to assume their initial response is incorrect and consider why that might be true. This is the method of prompting counterfactual thought used by Koonce (1992), Kennedy (1995), and Kadous et al. (2006). This active form of counterfactual thinking differs from the passive counterfactual mindset used in this study.

[11] The observation that the combination of "think outside of the box" instructions plus time pressure was necessary to result in revisions to the audit plan is a good example of what Klein (2013) describes as creative desperation, or insight coming to an individual in their time of need.

conscious attempts to do so fail (Galinsky et al. 2000). Consistent with literature in psychology in this area, I define a mindset as "a cluster of cognitive processes that are well-learned because they serve a functional purpose" (Galinsky et al. 2000, p256). Activation of these well-learned processes creates a change in cognitive processing style that carries over into subsequent, unrelated tasks (Smallman and Roese 2009).[12] For example, priming a counterfactual mindset helps mitigate the phenomenon of functional fixedness, or the tendency to focus on an object's most common use to the extent that other potential uses cannot be called to mind (Galinsky and Moskowitz 2000).[13]

Recall that auditors designing audit procedures for a familiar transaction will have difficulty due to constrained knowledge search resulting from strong cue-knowledge-idea links (Figure 1, Panel A). A familiar transaction brings to mind a specific set of audit evidence that the auditor has obtained successfully in the past and that suppress the consideration of further alternatives, similar to the problem of functional fixedness. Therefore, to the extent that a prompt facilitates the access of a wider range of task-related knowledge it will enable auditors to consider a more flexible set of potential audit evidence (Figure 2, Panel A).

[12] Consistent with the Galinsky et al. (2000) definition of mindset can relate to a variety of information processing styles that an individual acquires over time. A counterfactual mindset is just one of many mindsets used in the accounting literature. Griffith et al. (2014) study deliberative vs. implemental mindsets. While these two mindsets form a natural dichotomy, not all mindsets are related to each other or can be placed on a continuum in this manner. Other mindsets used in accounting are low-level and high-level construal mindsets from construal-level theory of psychological distance (Trope and Liberman 2010). Again, these mindsets form a dichotomy, but are unrelated to the counterfactual mindset, which is the topic of this study.

[13] The most famous of tasks illustrating this phenomenon, and the one chosen Galinsky and Moskowitz (2000), is the Duncker candle task. In this task, participants receive a box of thumbtacks and a candle along with instructions that they are to use these materials to affix the candle to the wall. To successfully complete this task, participants must remove the pushpins from their container, and realize that they can then use container itself as a stand for the candle. Participants were more likely to successfully complete this task after reading a story designed to elicit counterfactual thoughts and subsequently a counterfactual mindset. The story described a woman named Jane who attends a concert at which a member of the audience wins a free trip to Hawaii based on seat number. In the version of the story designed to elicit counterfactual thoughts Jane switches seats just prior to the announcement of the winning seat, but does not switch seats in the non-counterfactual conditions. In both conditions, Jane wins the trip to Hawaii.

In unfamiliar tasks it is less clear that priming a counterfactual mindset will be beneficial to idea generation since auditors need help focusing their knowledge search to relevant audit topics. Although it is possible that activating a wider range of knowledge is more likely to call to mind *any* knowledge that can be applied to the problem at hand, it seems more plausible that in a professional setting that an undirected broadening of activated knowledge will further interfere with the identification of *relevant* knowledge. For example, completion of an audit planning task requires the access of specific domain knowledge such as relevant accounting standards and auditing assertions (e.g., Vera-Muñoz et al. 2001). Therefore, I predict that in unfamiliar tasks that require specialized knowledge, prompting a counterfactual mindset will result in a broad, undirected search that is unlikely to yield knowledge that is applicable to the problem at hand (Figure 2, Panel A). I combine the differential benefits of priming a counterfactual mindset in the hypothesized interaction below (Figure 3):

> **H1**: A counterfactual mindset will improve the flexibility of audit evidence idea generation when a task is familiar, but will reduce or have no effect on audit evidence idea generation when a task is unfamiliar.

The Moderating Effect of Examples

Due to the prominence of recommended, or standardized, auditing procedures throughout audit firm and regulatory guidance, it is important to consider the robustness of the effects of a counterfactual mindset prime in this environment.[14] Research in engineering design shows that provision of an example can act as an anchor that limits the flexibility of ideas regarding other audit evidence that could be considered (Jansson and Smith 1991).[15] Several studies similarly

[14] The investigation of the joint effects of a counterfactual mindset prime and the provision of an example is closely related to recent calls for greater research on management control practices that consider the complementarity between controls and interventions acting as a system (Grabner and Moers 2013).

[15] The group brainstorming literature has documented several other detrimental effects of example provision, or idea sharing, on creativity and productivity (Dennis and Valacich 1993; Carpenter 2007; Lynch et al. 2009) that I believe are less concerning to an individual setting. One cause of these findings is production blocking, or cognitive

show that auditors anchor on prior year audit programs through reluctance to remove procedures from the audit program, but they do not hesitate to add additional work to the program indicating inherited programs are likely less sticky than self designs (Wright 1988; Biggs et al. 1988; Hammersley et al. 2011).[16] Therefore, to the extent that an example audit procedure is sufficiently different from the individual's own ideas it can be quite beneficial to the generation of further ideas (Nijstad et al. 2003). Below I will consider the joint effects of a counterfactual mindset prime and example provision on unfamiliar and familiar audit planning tasks respectively.

Unfamiliar Tasks

Recall that for unfamiliar tasks a counterfactual mindset will not help auditors narrow their knowledge search to relevant topics. When the task is unfamiliar the determination of what information is the most appropriate for that memory search can be very effort intensive (Nijstad et al. 2002; Nijstad and Stroebe 2006).[17] To the extent that the auditor can use an example as a search cue to more easily link the unfamiliar task to existing audit knowledge, the provision of an example will improve idea generation. Although an example has the potential to serve as a search cue in isolation, a counterfactual mindset will block the recognition of the example as a potential search cue.

---

interference, which occurs when outside ideas cause individuals to lose their train of thought and prematurely abort the cue-idea cycle (Diehl and Stroebe 1987; Stroebe et al. 2010).

[16] While these studies do not investigate the extent to which auditor program modifications are flexible, they do provide some assurance that auditors are unlikely to rely solely on the example without bringing their own experiences to bear on the task.

[17] Low (2004) finds that auditors with greater industry expertise are more likely to respond to changes in risks by changing procedures used in a prior year audit than are auditors with less expertise. These findings are consistent with SIAM in that experts can more easily process the task information into search cues and will subsequently retrieve more relevant knowledge, thus yielding benefits to idea generation. Low does not investigate the type of procedure changes so his study cannot speak to the creativity of the revised procedures directly, although his findings imply that industry experts would likely be able to produce a larger number of ideas, one component of divergent thought measured in the psychology literature, than non-experts (Guildford 1967).

In addition to increasing the consideration of alternatives, psychology research shows that counterfactual primes also improve the consideration of relationships between task-embedded cues (Kray et al. 2006). In an audit planning context this means that the counterfactual mindset will reinforce the link between the task and the example (i.e., emphasize the problem-idea link) such that it creates a fixation on the example as the answer, rather than as a search cue.[18] Research on counterfactual thinking and creativity in psychology finds that priming a counterfactual mindset can have detrimental, rather than beneficial or even neutral, effect on idea generation tasks containing examples (Kray et al. 2006).[19] However, unlike participants in those studies who possessed no expertise related to the task, auditors will eventually recognize the value of the example as a search cue and use it to access relevant knowledge, but delays in beginning this process will negate the benefits of example provision (Figure 2, Panel B). I make the following prediction regarding the joint effect of examples and counterfactual mindset for unfamiliar task (Figure 4, Panel A):

> **H2a:** For unfamiliar tasks, idea generation will be highest when an example is present and the auditor has a counterfactual mindset and least when an example is absent and the auditor does not have a counterfactual mindset or when the auditor has a counterfactual mindset, regardless of example provision.

Familiar Tasks

Recall that for familiar tasks a counterfactual mindset will help auditors widen their knowledge search beyond typical audit procedures learned through repetitive use. An example provided by firm or regulator guidance will expand auditors' knowledge search to the extent that any one particular example is sufficiently different enough from their own ideas to activate

[18] Focusing on relationships between task embedded information is also less cognitively taxing than engaging in the more difficult task of searching for relevant outside knowledge to apply (Nijstad and Stroebe 2006).

[19] The primary task used in this work was a product label idea generation task that always included the provision of an example response in the task materials and used psychology student subjects, who were likely unfamiliar with the field of advertising.

previously inaccessible knowledge (Nijstad et al. 2003). Therefore, the benefit of a counterfactual mindset prime provides the greater benefit to idea generation because it will activate more knowledge in all auditors, regardless of individual auditors' idiosyncratic standard responses, but an example will activate new knowledge only when sufficiently different from that standard response. Since auditors will not need to rely on the example as a search cue in familiar tasks, the benefits of the counterfactual mindset prime on idea generation will be robust to example presence (Figure 2, Panel B). I make the following prediction regarding the joint effect of examples and counterfactual mindset for familiar tasks (Figure 4, Panel B):

> **H2b**: For familiar tasks, idea generation will be highest when the auditor has a counterfactual mindset, regardless of example presence, lower when an example is present and the auditor does not have a counterfactual mindset, and least when an example is absent and the auditor does not have a counterfactual mindset.

IV. STUDY 2: EXPERIMENTAL METHOD

I test my hypotheses in a 2 x 2 x 2 between subjects experiment that presented participants with a hypothetical audit procedure design task. I manipulate *task* (higher familiarity versus lower familiarity), *counterfactual mindset* (counterfactual versus no counterfactual) and *example* (present versus absent). The participants were professional auditors with two to five years of experience and a rank of senior/in-charge or higher. Participants with this level of experience are appropriate for this task as they likely have experience designing or modifying audit procedures. In all conditions, participants completed a short thought exercise containing the counterfactual mindset manipulation before completing the primary task. The task required participants to read a short description of a hypothetical audit engagement and to list potential audit evidence items that they could obtain for the purposes of designing an audit program for a portion of that engagement. Figure 5 provides an outline of the experimental procedures. The author and a second independent researcher coded participant responses for four different attributes of idea generation: (1) the number of audit evidence ideas, (2) the breadth of audit evidence ideas, (3) the originality of the ideas relative to other participants, and (4) the degree of specification provided for each evidence idea.

Independent Variables

Task

The primary experimental task described for each participant one of two hypothetical audit engagements, manipulated between subjects, for which participants would be generating audit evidence ideas. The lower familiarity task was a case in which audit knowledge was applicable (i.e., the case contained audit *content*), but lacked a traditional auditing setting that participants could use to determine which portion of their knowledge would be most relevant

(i.e., the case lacked audit *context*) (Burns and Stalker 1961; Einhorn and Hogarth 1981; Bonner 1994). In this case participants identified audit evidence that could be obtained when auditing the number of golf balls that fit inside a school bus. The most relevant audit knowledge would likely be that related to auditing inventory accounts, however, the word inventory did not appear in the text of the case and no other details of the case were obviously inventory related. The higher familiarity task was a case where audit knowledge was applicable (i.e., the case contained audit *content*) and the setting provided clear contextual cues that participants could use to determine which portion of their knowledge would be most relevant (i.e., the case contained audit *context*) (Burns and Stalker 1961; Einhorn and Hogarth 1981; Bonner 1994). In this case participants identified audit evidence that could be obtained when auditing the revenue of a bridge tollbooth. The most relevant audit knowledge would be that related to auditing revenue accounts, which is quite salient due to the presence of the word revenue and other details of the case description. Both tasks are provided in their entirety in Figure 6, Panels B and C (see Appendix C for full experimental materials).

Counterfactual Mindset

In all conditions the instructions direct participants to list as many responses as they can, but the SIAM indicates that difficulties in retrieving diverse knowledge related to an individual search cue will be largely subconscious and prompts that again ask participants directly to consider alternatives to their initial response will likely be largely ineffectual. I prime a counterfactual mindset that will subconsciously influence subsequent cognition while participants complete the audit case. The counterfactual manipulation precedes the audit task and example manipulation for two reasons: (1) a mindset is by definition a state of mind that carries

over from a previous, unrelated task; and (2) a mindset influences information processing is therefore most influential when introduced prior to first consideration of the task.

Consistent with the related psychology literature, the mindset prime in the counterfactual condition asks participants to read a story designed to illicit counterfactual thoughts and to write down what thoughts the protagonist might be having (Figure 6, Panel A). In the story a woman named Jane attends a rock concert where a free trip to Hawaii is given to a member of the audience based on their seat number. In the counterfactual version of the story Jane switches seats just prior to winning the trip, thus Jane might be thinking thoughts such as, "If I had not switched seats, I would not have won the trip." The no counterfactual condition is identical to the counterfactual condition except that Jane does not switch seats prior to winning the trip. For a manipulation check, participants responded to a true or false question asking whether or not Jane switched seats in the story they read. Six (4.3%) participants failed this manipulation check with four of the six reporting that Jane had switched seats when she had not.[20]

Example Presence

I manipulate example provision by adding (omitting) a sentence at the end of the case materials that provides an example item of audit evidence and an example audit procedure that would incorporate that item of evidence. The example manipulation occurs after the counterfactual mindset manipulation so that it is processed as a part of the task and with the counterfactual mindset, when primed. The example also appears at the end of the case so that it has more meaning in the experimental context, since participants have no prior knowledge of the hypothetical audit client. In a real life audit setting an example might be presented before,

[20] The *counterfactual mindset* manipulation occurred before the audit task and is therefore unlikely to be influenced by the other two manipulated variables contained in that audit task. Accordingly, I observe a distribution of counterfactual mindset manipulation check failures that do not statistically differ across cases (higher familiarity = 4.3%; lower familiarity = 4.3%; $Z < 0.01$; $p = 1.000$, two-tailed) and example conditions (example = 1.5%; no example = 7.1%; $Z = 1.75$; $p = 0.102$, two-tailed).

during, or after the review of critical task details. However, as long as the example is considered with the same mindset as the other task information there is no theoretical reason to believe that the two manipulated factors will interact differently that the hypothesized pattern of results.

The example provided for the school bus (lower familiarity) case was to obtain the dimensions of the school bus for use in an analytical procedure that would estimate how many golf balls could fit in the bus. The example provided for the tollbooth (higher familiarity) case was to obtain the average daily traffic flow across the bridge for use in an analytical procedure to estimate the expected annual revenues. In the no example condition the sentence containing the example was omitted from the case.

The nature of the example will, of course, influence the effectiveness of this manipulation. I select a fairly common, non-account specific example, which I believe biases against finding results since more common examples are more likely to have already been considered by participants. Providing a less typical example, theory predicts, would likely amplify the predicted effects, but not change the overall shape of the interaction (Njistad et al. 2003). The example manipulations are included in Figure 5, Panels B and C. For a manipulation check participants responded to a true or false question containing the example audit evidence that asked whether that sentence appeared in their case materials. Twenty-three (16.7%) participants failed the manipulation check with 13 of the 23 reporting that they had seen the example when they had not.[21]

Dependent Variables

My primary dependent variable is a composite measure of the quantity and quality of audit evidence ideas generated that is based upon the scoring system utilized by Guilford's

[21] I do not observe an effect of case (higher familiarity = 17.4%; lower familiarity = 15.9%; Z = 0.23; p = 0.819, two-tailed) or counterfactual mindset (counterfactual = 17.6%; no counterfactual = 15.7%; Z = -0.31; p = 0.761, two-tailed) on whether or not participants failed the *example* manipulation check question.

Alternate Uses Task (1967). The alternative uses task asks participants to name various possible uses for a common object such as a brick or a paperclip and scoring of responses evaluates four different dimensions: fluency, breadth elaboration, and originality.[22] The author and an independent second coder, who were both blind to the experimental condition, coded each audit evidence idea for these four criteria. Prior literature in creativity or divergent thinking typically assumes that these components together measure a single construct (e.g., Karakelle 2009; Lewis and Lovatt 2013). The SIAM literature indicates that while taken together these measures should indicate better idea generation overall, each measure individually represents a more specific aspect of the idea generation construct on its own. Fluency measures productivity through volume of ideas; breadth measures content through diversity of ideas; and elaboration and originality measure quality through specificity and uniqueness of ideas respectively (Stroebe et al. 2010).

While these attributes of idea generation are well established in the creativity and brainstorming literature, it is not necessarily obvious how they equate to benefits in audit program design so I consider each in turn. The objective of audit planning is to ultimately design audit procedures that maximize the persuasiveness, and thus sufficiency, of evidence collected while also maximizing efficiency to the greatest extent possible. Auditors with a greater number and variety of audit evidence items to chose from should be more successful in this endeavor. Additionally, choosing new items or having a greater number of evidence items to alternate between will help auditors provide unpredictability in procedures. Therefore, fluency and breadth are relevant to the audit setting. In audit procedure design, specificity is very important, particularly when auditing judgmental areas. For example, a bank statement obtained from the

[22] Guildford's Alternate Uses Test (1967) refers to breadth as flexibility and the SIAM literature refers to breadth as idea diversity (Stroebe et al. 2010).

auditee is less persuasive than a bank statement obtained directly from the bank. If no source were specified in the audit program then the effectiveness of the work step could be jeopardized. Therefore, elaboration is relevant to the audit setting. However, while original ideas are necessary for big picture process improvements it is not the case creating audit procedures from a set of highly original audit evidence, that omits unoriginal evidence such as bank statements, will lead to better audit procedures.

I perform a factor analysis on these four dimensions and find that they in fact load on two different factors with eigenvalues of greater than one. Three components (factor loading): fluency (.93), breadth (.75), and elaboration (.85) load on the first factor, which I refer to as idea generation. The remaining component of originality loads in isolation on the second factor. The idea generation factor contains the components that are most critical to day-to-day success of audit procedure design in the ability to generate volume, breath and specificity of ideas. Originality, by contrast, is more critical to the evolution of new methods of auditing and future process improvements. I perform tests of my hypotheses on both a composite measure representing the idea generation factor as well as the originality factor (see Figures 7 and 8 for example responses). I describe each of the four measures individually in greater detail below.

Fluency

Fluency measures the quantity of evidence ideas that a participant identifies. Each evidence idea listed by the participant that is distinctly different from the other evidence ideas listed by that same participant received one point. Nonsensical responses or responses that referred to parts of the audit that were outside the scope of the experimental task received no points. The final fluency score for each participant is the sum of all points assigned for evidence

ideas identified. The coder agreement for the fluency measure was 89.0% and all disagreements were mutually resolved.

Breadth

Breadth measures the variety of evidence types included in a participant's idea set. I define the different categories of possible audit evidence based on the types of audit evidence listed in the AICPA auditing standards (AU section 326). These categories include inspection of documents, inspection of tangible items, observation, inquiry, confirmation, recalculation, reperformance, analytical procedures and scanning. The two coders mapped each evidence idea provided by participants to one of these nine categories. Participants received one point for each category in which they had listed at least one evidence idea. The final breadth score for each participant is the sum of all categories of audit evidence utilized across their responses. The coder agreement for the breadth measure was 86.1% and all disagreements were mutually resolved.

Elaboration

The elaboration dimension measures the amount of detail that the participant includes in their response. The coders assigned one point per modifier or additional detail provided that elaborates on the idea generated. For example, in the lower familiarity (school bus) case if the participant simply identified the school bus itself as an item of audit evidence that would receive no elaboration points. However, if the participant identified the school bus itself and elaborated that the purpose of examining this evidence was to measuring the interior dimensions of the bus then the response would receive one point of elaboration. The final elaboration score is the sum

of all elaboration points awarded across all items of evidence identified. The coder agreement for the elaboration measure was 76.6% and all disagreements were mutually resolved.[23]

Originality

Originality measures the uniqueness of the response relative to the entire pool of participants such that rare responses receive a higher score than common responses. Since responses were often case specific, comparisons are made only to other participants in the same familiarity condition. Specifically, I assign points to responses that were identified by fewer than 15% or fewer respondents. Since approximately 70 participants completed each case this results in participants receiving at least one point for any response indicated by ten or fewer other participants. The scoring system provided more points to more uncommon responses such that responses given by only one participant received ten points; responses given by only two participants received nine points, etc. The final originality score for each participant is the sum of all points assigned for responses indicated by 15% or fewer respondents and divided by the participants fluency scores.[24] Scaling by fluency was necessary to prevent contamination of the measure based on response volume. For example, a participant who gave ten responses each earning a single originality point should not be considered equally original to a participant that gave a single, truly unique response. The originality measure was computed based off of the evidence idea names assigned by the coders and used to compute the fluency measure, no further coding was performed specifically for this measure.

[23] The majority of the disagreements (88.0%) represent coder responses within a range of plus or minus one of each other and the simple correlation between raters was 0.72.

[24] Results are robust to alternative cut-off points for receiving originality points, such as awarding points only to responses given by 5% or fewer of respondents.

V. STUDY 2: RESULTS

Participants

Participants in this study were 138 experienced audit professionals from two Big 4 audit firms recruited during firm organized meetings. These subjects all held the rank of senior/in-charge and had on average (standard deviation) 3.2 (1.2) years of auditing experience in a wide range of client industries. Posttest questions inquiring of the frequency with which participants encountered situations requiring audit program modifications and how much experience they had in such situations indicate that the participants possessed adequate professional experience to complete the experimental task. Participates indicated that on average (standard deviation), 6.2 (1.9) on a nine-point scale (1 = Very little extent; 9 = Very great extent), they agreed with the statement, "I often encounter situations in which I must modify an existing audit program." They also reported possessing at least moderate experience in designing and modifying audit procedures, rating their experience on average (standard deviation) at 5.8 (1.9) also on a nine point scale (1 = No experience; 9 = Extensive experience). Finally, on average (standard deviation) participants reported that they believed they were able to add value to the audit in the past when they were required to design or modify audit procedures.[25] Overall, these measures indicate that participants had sufficient professional and task specific experience.

Test of Hypotheses

Joint Effect of Counterfactual Mindset and Task

Participants' idea generation scores are tabulated in Table 13 and illustrated graphically in Figures 9 and 10. The idea generation scores provided in Table 13 reflect the factor composite score that combines the fluency, breadth, and elaboration measures. I include time spent to

[25] Participants rated their belief that they were able to add value to the audit through audit program modification on a nine-point scale where the end points were labeled 1 = Very little extent and 9 = Very great extent as an average (standard deviation) of 6.8 (1.4).

complete the task as a covariate in the ANCOVA model to control for effort that the participant exerted on the task (Table 14, Panel A). None of the experimental manipulations individually or jointly had a significant impact on the amount of time participants took to complete the task. Recall that H1 predicts that a counterfactual mindset will improve audit evidence idea generation when familiarity is higher, but not when familiarity is lower (Figure 3). I find a significant interaction of counterfactual mindset and familiarity ($F_{1,129} = 5.56$; $p = 0.010$, one-tailed) (shown graphically as Figure 9). Follow-up simple effects tests show a significant positive effect of counterfactual mindset given higher familiarity ($F_{1, 129} = 4.66$; $p = 0.017$, one-tailed) and a non-significant negative effect of counterfactual mindset given lower familiarity ($F_{1, 129} = 1.42$; $p = 0.236$, two-tailed) (Table 14, Panel B).[26] Therefore, H1 is supported for the overall idea generation measure.

Originality scores are provided in Table 15. The disordinal interaction of familiarity and counterfactual mindset is not significant ($F_{1, 129} = 0.95$; $p = 0.331$, two-tailed) (Table 16, Panel A). Follow-up simple effects tests show a non-significant negative effect of counterfactual mindset given higher familiarity ($F_{1, 129} = 1.48$; $p = 0.226$, two-tailed) and a non-significant negative effect of counterfactual mindset given lower familiarity ($F_{1, 129} = 0.03$; $p = 0.863$, two-tailed) (Table 16, Panel B). Therefore, H1 is not supported for the originality measure.

Overall, I find that priming a counterfactual mindset improves overall idea generation in familiar settings, but does not influence the originality of those ideas generated. This likely due to the fact that having a highly original set of audit evidence ideas is not necessarily indicative of

[26] These results are robust to the inclusion of audit firm as a factor in the ANCOVA model. Interestingly, I find an unpredicted three-way interaction of firm, familiarity and example ($F_{1,129} = 4.47$; $p = 0.036$, two-tailed) in which the example provided the greatest benefit to auditors from Firm 2 in the less familiar case. This is likely because employees of the same audit firm will have been exposed to more similar standard procedures than will auditors from different firms. Since an example is most beneficial when it is dissimilar from previous experiences, I conjecture that the particular example used in the less familiar task was more dissimilar to standard procedures employed by Firm 2 than by Firm 1.

a higher quality idea generation, as is assumed in the idea generation literature as a whole. For example, some forms of very common audit evidence such as third party confirmations are likely critical to most audit engagements and ignoring this type of evidence in favor of something more original, unless it provided equally strong assurance of existence, could impair audit quality. I, therefore, conclude that H1 is supported.

Joint Effect of Counterfactual Thinking and Example Provision

H2 predicts a three-way interaction that is comprised of two two-way ordinal interactions of counterfactual mindset and example provision that are moderated by task (Figure 4). Since the traditional ANOVA three-way interaction term tests for disordinal interactions it is not a good test of my predictions. Additionally, my theory does not allow me to predict a directional main effect of task familiarity such that I can contrast code the full three-way interaction. Instead, I test the two ordinal interactions individually using contrast codes representing the specific pattern of predicted results (Buckless and Ravenscroft 1990). H2a predicts that when familiarity is lower an example will benefit idea generation greatly, but priming a counterfactual mindset will negate these benefits in addition to providing no benefit when prompted in isolation (Figure 4, Panel A). H2b predicts that when familiarity is higher an example will benefit idea generation moderately, but priming a counterfactual mindset will provide the greatest benefit regardless of example presence (Figure 4, Panel B).

When familiarity is lower (H2a), I find a significant interaction of counterfactual mindset and example for idea generation as presented in Table 14, Panel C and illustrated in Figure 10, Panel A ($F_{1, 129} = 50.96$; $p < 0.001$, one-tailed).[27] Simple effects follow up tests (not tabulated) show a marginally significant benefit of example provision in the absence of a counterfactual

[27] An alternative, also theory consistent, contrast codes of absent/no counterfactual (+1), present/no counterfactual (+3), absent/counterfactual (-2), and present/counterfactual (-2) are also significant ($F_{1, 129} = 35.86$; $p < 0.001$, one-tailed).

mindset ($F_{1, 129}$ = 1.66; p = 0.100, one-tailed) and a marginally significant negative effect of counterfactual mindset in the presence of an example ($F_{1, 129}$ = 1.87; p = 0.087, one-tailed). The significant ordinal interaction and weakly supporting simple effects support the predictions of H2a for the overall idea generation dependent variable. However, the originality variable does not display the predicted pattern of results and instead suggests a (non-significant) positive effect of any prompts individually or in combination (Table 16, Panel C). Therefore, in less familiar tasks the provision of an example is beneficial to overall idea generation, but not originality and only when the idea generator is not in a counterfactual mindset.

When familiarity is higher (H2b), I find a significant interaction of counterfactual mindset and example for idea generation as presented in Table 14, Panel C and illustrated in Figure 10, Panel B ($F_{1, 129}$ = 22.08; p < 0.001, one-tailed).[28] Simple effects follow up tests (not tabulated) show a significant effect of counterfactual mindset in the absence of an example ($F_{1, 129}$ = 3.63; p = 0.030, one-tailed), but non-significant effects of counterfactual mindset in the presence of an example ($F_{1, 129}$ = 1.32; p = 0.127, one-tailed) and example in the absence of counterfactual ($F_{1, 129}$ = 1.17; p = 0.141, one-tailed).[29] The significant ordinal interaction and supporting simple effect of counterfactual in the absence of example support the predictions of H2b for the overall idea generation dependent variable. However, the originality variable does not display the predicted pattern of results and instead suggests a (non-significant) detrimental effect of a counterfactual mindset on originality in more familiar settings (Table 16, Panel C). Therefore, in more familiar tasks the counterfactual mindset prompt is beneficial to overall idea generation, but not originality, regardless of example presence.

[28] An alternative, also theory consistent, contrast codes of absent/no counterfactual (-3), present/no counterfactual (+1), absent/counterfactual (+1), and present/counterfactual (+1) are also significant (F1, 129 = 4.64; p = 0.017, one-tailed).

[29] These results, although weaker in the simple effects tests, are robust to firm effects.

Overall, I find results consistent with my predictions for general idea generation, but not the originality of those ideas generated. As discussed previously, this likely due to the originality measure not being indicative of idea set quality when it suppresses more routine, yet desirable, ideas. I, therefore, conclude that both H2a and H2b are supported.

Supplemental Analysis Component Measures

The primary dependent measure is a factor composite score that combines three coded components: fluency, breadth, and elaboration. Although I do not make predictions regarding these component measures I analyze them individually for consistency with my predictions and indications of whether one or more of the measures drive the observed idea generation results. I find that fluency (i.e., the number of ideas) and elaboration (i.e., the level of idea specification), which are the only two of the dimensions coded that do not compare participants against an outside standard (i.e., audit evidence categories and other participant responses) provide the most similar patterns of results both to each other and to the composite measure. However, it appears that the breadth measure is the biggest driver of the benefits of example provision. I provide a more detailed analysis of the individual measures below.

Fluency

Fluency is the component measure representing the number of ideas generated by the participant. Descriptive statistics for fluency are presented in Table 17, Panel A and the ANCOVA and simple effects tests are presented in Table 18. The interaction of familiarity and counterfactual mindset is significant ($F_{1, 129} = 3.84$; $p = 0.026$, one-tailed). The simple effects follow-up tests indicate a marginally significant positive effect of counterfactual mindset given higher familiarity ($F_{1, 129} = 2.66$; $p = 0.053$, one-tailed) and a non-significant negative effect of counterfactual mindset given lower familiarity ($F_{1, 129} = 1.33$; $p = 0.252$, two-tailed). These

results are consistent with the predictions of H1 indicating that the number of ideas generated is a strong driver of the overall idea generation results. For H2 the results are not as strong as those for H1. Although both the lower (H2a) and higher (H2b) familiarity conditions display the hypothesized pattern of results, the follow up simple effects test are not significant indicating that the number of ideas generated is not a strong driver of the overall idea generation results for H2.

Breadth

Breadth is the component measure representing the variety of ideas generated by each participant. Descriptive statistics for breadth are presented in Table 17, Panel B and the ANCOVA and simple effects tests are presented in Table 19. The interaction of familiarity and counterfactual mindset is not significant ($F_{1, 129} = 0.86$; $p = 0.179$, one-tailed). These results are inconsistent with the predictions of H1 indicating that the breadth of ideas generated is not a strong driver of the overall idea generation results.

For H2 the results are mixed. When familiarity is lower (H2a), I find a significant planned contrast interaction of counterfactual mindset and ($F_{1, 129} = 59.54$; $p < 0.001$, one-tailed). The simple effect follow up tests indicate a significant positive effect of example in the absence of a counterfactual mindset ($F_{1, 129} = 3.89$; $p < 0.026$, one-tailed). When familiarity is high (H2b), however, the overall pattern shows no benefit of example despite finding a significant planned contrast interaction of counterfactual mindset and example ($F_{1, 129} = 9.19$; $p = 0.003$, one-tailed) and the simple effects follow up tests are not significant. These results indicate that the breadth of ideas generated is a strong driver of the overall idea generation results for H2a (lower familiarity) but not for H2b (higher familiarity).

Elaboration

Elaboration is the component measure representing the level of specification that the participant provides regarding the ideas generated. Descriptive statistics for elaboration are presented in Table 17, Panel C and the ANCOVA and simple effects tests are presented in Table 20. The interaction of familiarity and counterfactual mindset is significant ($F_{1, 129} = 6.54$; $p = 0.006$, one-tailed).[30] These results are consistent with the predictions of H1 indicating that the specificity of ideas generated is a strong driver of the overall idea generation results.

For H2, however the results are mixed. When familiarity is lower (H2a) I find a significant planned contrast interaction of counterfactual mindset and ($F_{1, 129} = 23.34$; $p < 0.001$, one-tailed), but the simple effect follow up tests are not significant. When familiarity is higher (H2b), however, I find a significant planned contrast interaction of counterfactual mindset and example ($F_{1, 129} = 27.84$; $p < 0.001$, one-tailed). The simple effects follow up tests indicate a significant positive effect of a counterfactual mindset in the absence of an example ($F_{1, 129} = 8.07$; $p = 0.003$, one-tailed) and example provision in the absence of a counterfactual mindset ($F_{1, 129} = 4.99$; $p < 0.014$, one-tailed). These results indicate that the specificity of ideas generated is a strong driver of the overall idea generation results for H2b (higher familiarity) but not for H2a (lower familiarity).

Supplemental Analysis Regarding Audit Efficiency

One of the primary motivations for this study is that promoting flexible idea generation will increase the likelihood of discovering audit process improvements. I find evidence that priming a counterfactual mindset and providing an example procedure, both individually or

[30] The simple effects follow-up tests indicate a significant positive effect of counterfactual mindset given higher familiarity ($F_{1,129} = 6.50$; $p = 0.006$, one-tailed) and a non-significant negative effect of counterfactual mindset given lower familiarity ($F_{1,129} = 1.18$; $p = 0.280$, two-tailed).

jointly, improve audit efficiency for familiar tasks (ANOVA results not tabulated). Following the primary audit evidence idea generation task participants created an audit program from the procedures listed and estimated the number of hours that would be required to complete their program. Compared to the no counterfactual mindset and no example condition participants reported that their audit plans would take significantly fewer hours to complete when they received an example ($F_{1,\ 122} = 3.92$; $p = 0.050$, two-tailed), when they were primed with a counterfactual mindset ($F_{1,\ 122} = 4.16$; $p = 0.044$, two-tailed), and when they both received an example and were primed with a counterfactual mindset ($F_{1,\ 122} = 4.13$; $p = 0.044$, two-tailed). These results are robust to controlling for the participants' judgments of how much assurance their programs provided. Therefore, it appears that participants believed that they were able to form more efficient audit programs when prompted to engage in flexible idea generation.

VI. CONCLUSIONS

I examine the potential benefits of counterfactual mindset and example provision on promoting flexible auditor thinking in audit planning design. Theory indicates that priming a counterfactual mindset helps auditors activate a wider range of knowledge when generating audit evidence ideas for a familiar auditing context, regardless of example presence. In unfamiliar auditing contexts, however, the introduction of a counterfactual mindset negates the benefits of example provision. While the majority of auditing literature to date demonstrates how counterfactual thinking can be beneficial to auditor judgment, I find that priming a counterfactual mindset can have negative effects for unfamiliar audit planning tasks. Overall, my theory and findings indicate that when auditors desire increased flexibility in idea generation for the redesign of familiar transactions, priming a counterfactual mindset is the more beneficial prompt. If auditors desire increased flexibility in idea generation when designing audit procedures for unfamiliar transactions the provision of an example is the more beneficial prompt. Finally, in no setting did the joint use of both prompts together provide additive benefits to idea generation.

Practical Implications

Although the design of this study purposely abstracts away from a real audit setting in order to provide a strong test of theory, I believe that the concepts could be fairly easily implemented in practice. In an effort to increase standardization, the major accounting firms have invested in technology that guides auditors through the firm's methodology. This technology could also be used to promote flexible idea generation without sacrificing control. Audit seniors often take the first pass at setting up a reoccurring engagement. When doing so, the technology system could ask, "Are you planning for a new transaction/area or for a transaction/area audited in prior year?" If the auditor response it is a recurring transaction then

the next question would be one that evokes a counterfactual mindset (e.g., take a moment to list anything you believe we should have done differently in prior year). If the auditor responds it is a new transaction then they would not receive the counterfactual question and would be directed to seek out an example before continuing. In both cases, the system could maintain a database of potential evidence to be utilized during formal stages of planning and validated during walk through procedures.

In practice it may be difficult to determine an individual's familiarity level with an audit area or transaction. For example, recall that in Study 1 I find evidence indicative of auditors with greater experience being more likely to misperceive relatively *unfamiliar* transactions as relatively *familiar*. Future research should examine individual ability to assess familiarity. If, in the interim, audit firms wished to enact the recommendations made in this study, they could do so by utilizing the counterfactual prompt only for less experienced auditors. The highest risk scenarios for an audit firm are those that are unfamiliar to their most experienced auditors. In these cases a counterfactual mindset negates the benefits of example provision and thus should be avoided. However, it is important to note that prompting a counterfactual mindset is relatively risk free since in no condition did any of the auditors perform worse than they did in the absence of the manipulated factors (i.e., no counterfactual/example absent condition). Less experienced auditors appear to be more aware about what is or is not familiar to them and work in less risky areas (i.e., less judgmental and more familiar to the firm as a whole). These areas could benefit from a counterfactual mindset, as it is the stronger stimulant of ideas in familiar tasks. Since the senior/in-charge auditors typically take the first pass at audit planning and are most involved with using audit technology interfaces, the recommended implementation described previously could be a viable implementation.

Limitations

This study has several limitations that indicate interesting avenues for future research. First, the effect of an example on idea generation is somewhat dependent on the example selected. The examples used in this study were fairly simple and common audit evidence, which biases against finding the predicted results. An example that is less usual is likely to amplify the effects I predict, but there is also the possibility that a boundary condition exists at which an example that is too unusual will be dismissed as irrelevant by the idea generator. Additionally, as a transaction increases in uniqueness, or firm level unfamiliarity, it becomes more difficult to obtain an example. One encouraging item to note here that the lack of ability to identify an example could become a critical red flag to audit firms that consultation with regulatory agencies may be necessary to identify an appropriate audit approach or consider resignation of the client if one cannot be identified. Future research should systematically vary features of the example to determine the most effective form for promoting flexible idea generation.

Second, in my operationalization of counterfactual mindset I use a manipulation that prompts a specific form of mental simulation and that I believe biases against finding my predicted results. However, the psychology literature has identified various attributes of mental simulations that change the effect on individuals' cognition and their subsequent actions. (1) Simulations may be either upward or downward in nature. Upward simulations create a scenario that is more desirable than reality while downward simulations create a scenario that is less desirable than reality (Wong et al. 2009). (2) Mental simulations may be counterfactual or prefactual. A counterfactual simulation involves modifying the facts of what actually happened in order to create a new, hypothetical scenario. Alternatively, prefactual simulations are fully fictional as they are constructed entirely from the individual's imagination of what may occur in

the future (Van Boven et al. 2009). (3) As individuals build simulations they may add or subtract various components in order to test the effects of doing so in their mental laboratory. Whether the nature of a mental simulation overall is more additive or subtractive can influence the cognitive processing style employed during subsequent activities (Van Boven et al. 2009; Wong et al. 2009).[31] (4) Finally, the simulation can vary in magnitude, or degree of departure from reality. The greater the magnitude, the less believable the simulation, and the less impact it has on future cognition and actions (Kahneman and Tversky 1982). My experiment uses an upward, additive, counterfactual prompt with a relatively small magnitude. Future research may wish to investigate the use of downward, subtractive, prefactual prompts and/or different combinations of these attributes with varying magnitudes.

Third, in designing a strong manipulation of familiarity it was beneficial to change many features of the task. Since I was unaware of the specific backgrounds of the auditor participants, it was not possible to know what types of tasks they were and were not familiar with *a priori*. Therefore, the design of the lower familiarity task used a case that was almost certainly equally unfamiliar to all participants. To the extent that a specific, subtler part of this manipulation drove the familiarity results, I am unable to tease this out in my results.

Finally, I cannot unilaterally recommend the interventions examined here for implementation in all audit settings, as other variables may exist in the real world that would interact in an undesirable way. Future research should investigate whether constructs such as expertise, accountability, information load, information relevance, incentives, and feedback interact differently with a counterfactual mindset and example provision than the relationships documented in this study.

[31] Cognitive processing style refers to the way that an individual accumulates information from their surroundings, organizes that information and subsequently uses it to make a decision or judgment (Leonard and Scholl 1999).

Despite these limitations, this is the first study I am aware of that explicitly examines flexibility of audit procedure design and more specifically investigates ways to enhance that idea generation process. Introduction of the counterfactual mindset to the accounting literature extends research that employs counterfactual thinking to improve auditor judgments to areas where conscious prompts to consider alternatives previously have failed (e.g., Heiman 1990). Additionally, I provide evidence of two factors, example provision and familiarity, that moderate the effect of counterfactual mindset in idea generation tasks. Example presence and familiarity may explain the documented negative effects of counterfactual mindset on creative generation tasks in the psychology literature (Kray et al. 2006). The effects of familiarity emphasize the need for further research on when the effects of counterfactual thinking research, performed largely with student subjects performing trivial tasks, generalize to professional subjects performing tasks within their area of expertise.

REFERENCES

Alissa, W., Capkun, V., Jeanjean, T., & Suca, N. (2014). An empirical investigation of the impact of audit and auditor characteristics on auditor performance. *Accounting, Organizations and Society*, 1–16.

American Institute of Certified Public Accountants (AICPA). (1989). Auditing Accounting Estimates.

American Institute of Certified Public Accountants (AICPA). (2006). Performing Audit Procedures in Response to Assessed Risks and Evaluating the Audit Evidence Obtained.

American Institute of Certified Public Accountants (AICPA). 2006. *AU section 326: Audit Evidence*. New York: AICPA

American Institute of Certified Public Accountants (AICPA). (2010). Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement (Redrafted), Statements on Auditing Standards 1–53.

Amabile, T. M. (1996). *Creativity In Context: Update To The Social Psychology Of Creativity.* Westview Press: Boulder.

Anonymous Audit Executive Director. [Executive Director at Big 4 Accounting Firm.] Interview. 2012.

Anonymous Audit Senior. [Senior Staff Member at Mid-Sized Accounting Firm.] Interview. 2012.

Arning, R. F. "Achieving Success in the Current Environment." The University of Illinois at Urbana-Champaign. Business Instructional Facility, Champaign, IL. 10 October 2013.

Asare, S. K., Trompeter, G. M., & Wright, A. M. (2000). The Effect of Accountability and Time Budgets on Auditors' Testing Strategies. *Contemporary Accounting Research*, *17*(4), 539–560.

Asare, S. K., & Wright, A. M. (2004). The Effectiveness of Alternative Risk Assessment and Program Planning Tools in a Fraud Setting. *Contemporary Accounting Research*, *21*(2), 325–352.

Bedard, J. C., & Biggs, S. F. (1991). The effect of domain-specific experience on evaluation of management representations in analytical procedures. *Auditing: a Journal of Practice & Theory,* 10, 77–90.

Biggs, S. F., Mock, T. J., & Watkins, P. R. (1988). Auditor's use of analytical review in audit program design. *The Accounting Review*, *63*(1), 148–161.

Bonner, S. E. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society*, *19*(3), 213–234.

Bonner, S. E. (2008). *Judgment and decision making in accounting*. Pearson Education, Inc.: Upper Saddle River, New Jersey.

Bowlin, K. (2011). Risk-based auditing, strategic prompts, and auditor sensitivity to the strategic risk of fraud. *The Accounting Review* 86 (4): 1231-1253.

Brazel, J. F., Agoglia, C. P., & Hatfield, R. C. (2004). Electronic versus face-to-face review: The effects of alternative forms of review on auditors' performance. *The Accounting Review*, *79*(4), 949-966.

Budescu, D. V., Peecher, M. E., & Solomon, I. (2012). The joint influence of the extent and nature of audit evidence, materiality thresholds, and misstatement type on achieved audit risk. *Auditing: A Journal of Practice & Theory*, *31*(2), 19–41.

Burns, T. & Stalker, G. M. (1961). *The Management of Innovation.* Oxford University Press: Oxford.

Cardoso, C., & Badke-Schaub, P. (2011). Fixation or Inspiration: Creative Problem Solving in Design. *The Journal of Creative Behavior*, *45*(2), 77–82.

Carpenter, T. D. (2007). Audit Team Brainstorming, Fraud Risk Identification, and Fraud Risk Assessment: Implications of SAS No. 99. *The Accounting Review*, *82*(5), 1119–1140.

Chen, C. X., Trotman, K. T., & Zhou, F. H. (2014). Nominal vs. Interacting Electronic Fraud Brainstorming in Hierarchical Audit Teams. *The Accounting Review*.

Chen, C. X., Williamson, M. G., & Zhou, F. H. (2012). Reward System Design and Group Creativity: An Experimental Investigation. *The Accounting Review*, *87*(6), 1885–1911.

Chipman, S. M. The American Accounting Association Annual Meeting. Anaheim, CA. 5 August 2013.

Church, B. K., & Shefchik, L. B. (2012). PCAOB Inspections and Large Accounting Firms. *Accounting Horizons*, *26*(1), 43–63.

Dennis, A. R., & Valacich, J. S. (1993). Computer brainstorms: More heads are better than one. *Journal of applied psychology*, *78*(4), 531.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology*, *53*(3), 497.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual review of psychology*, *32*(1), 53-88.

Fairchild, J., Cassidy, S., & Cushenbery, L. (2011). Integrating Technology with the Creative Design Process. In A. Mesquita, *Technology for Creativity and Innovation Tools, Techniques and Applications*. Information Science Reference: Hershey.

Galinsky, A. D., & Kray, L. J. (2004). From thinking about what might have been to sharing what we know: The effects of counterfactual mind-sets on information sharing in groups. *Journal of Experimental Social Psychology*, *40*(5), 606–618.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as Behavioral Primes: Priming the Simulation Heuristic and Consideration of Alternatives. *Journal of Experimental Social Psychology*, *36*(4), 384–409.

Galinsky, A. D., Moskowitz, G. B., & Skurnik, I. (2000). Counterfactuals as self-generated primes: The effect of prior counterfactual activation on person perception judgments. *Social Cognition*, *18*(3), 252–280.

Gibbins, M. (1984). Propositions about the psychology of professional judgment in public accounting. *Journal of Accounting Research*, *22*(1), 103–125.

Gibbins, M., & Newton, J. D. (1994). An empirical exploration of complex accountability in public accounting. *Journal of Accounting Research*, *32*(2), 165–186.

Gibbins, M., Salterio, S. E., & Webb, A. (2001). Evidence about auditor–client management negotiation concerning client's financial reporting. *Journal of Accounting Research*, *39*(3), 535–563.

Grabner, I., & Moers, F. (2013). Management control as a system or a package? Conceptual and empirical issues. *Accounting, Organizations and Society*, *38*(6), 407-419.

Griffith, E. E., J. S. Hammersley, and K. Kadous. 2014. Auditing complex estimates: Understanding the process used and problems encountered. *Contemporary Accounting Research*, forthcoming.

Guilford, J. P. (1967). *The Nature of Human Intelligence.* McGraw-Hill: New York.

Hammersley, J. S. (2011). A review and model of auditor judgments in fraud-related planning tasks. *Auditing: A Journal of Practice & Theory*, *30*(4), 101–128.

Hammersley, J. S., Johnstone, K. M., & Kadous, K. (2011). How do audit seniors respond to heightened fraud risk? *Auditing: A Journal of Practice & Theory*, *30*(3), 81–101.

Heiman, V. B. (1990). Auditors' assessments of the likelihood of error explanations in analytical review. *The Accounting Review*, *65*(4), 875–890.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61*(1), 569–598.

Herron, E. T. (2012). *Characteristics of Creativity in Relation to Auditors' Recognition of Fraud Cues and Response to Perceived Fraud Risk*. Oklahoma State University.

Hoffman, V. B., & Zimbelman, M. F. (2009). Do Strategic Reasoning and Brainstorming Help Auditors Change Their Standard Audit Procedures in Response to Fraud Risk? *The Accounting Review*.

International Federation of Accountants (IFAC). (2009). *International Standard on Auditing 240: The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements. Handbook of International Auditing, Assurance, and Ethics Pronouncements*. New York: IFAC.

Jamal, K., & Tan, H.-T. (2010). Joint Effects of Principles-Based versus Rules-Based Standards and Auditor Type in Constraining Financial Managers' Aggressive Reporting. *The Accounting Review*, *85*(4), 1325–1346.

Joyce, E. J. (1976). Expert judgment in audit program planning. *Journal of Accounting Research*, *14*, 29–60.

Joyce, E. J., & Biddle, G. C. (1981). Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research*, *19*(1), 120–145.

Kachelmeier, S. J., & Messier, W. F., Jr. (1990). An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *The Accounting Review*, 209–226.

Kachelmeier, S. J., Reichert, B. E., & Williamson, M. G. (2008). Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research*, *46*(2), 341–373.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman & P. Slovic, *Judgment under uncertainty: Heuristic and biases*, 201–211.

Karakelle, S. (2009). Enhancing fluent and flexible thinking through the creative drama process. *Thinking Skills and Creativity*, 4(2), 124–129.

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, 31(2), 231–245.

Kennedy, J. (1995). Debiasing the curse of knowledge in audit judgment. *The Accounting Review*, 249–273.

Kennedy, J., Kleinmuntz, D. N., & Peecher, M. E. (1997). Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research*, 35(Supplement), 105-123.

Klein, G. (2013). *Seeing What Others Don't: The Remarkable Ways We Gain Insight.* PublicAffairs: New York.

Koonce, L. (1992). Explanation and counterexplanation during audit analytical review. *The Accounting Review*, *67*(1), 59–76.

Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, *91*(1), 69–81.

Kray, L. J., Galinsky, A. D., & Wong, E. M. (2006). Thinking within the box: The relational processing style elicited by counterfactual mind-sets. *Journal of Personality and Social Psychology*, *91*(1), 33–48.

Leonard, N. H., & Scholl, R. W. (1999). Information processing style and decision making. *Journal of Organizational Behavior*, *20*, 407–420.

Lewis, C., & Lovatt, P. J. (2013). Breaking away from set patterns of thinking: Improvisation and divergent thinking. *Thinking Skills and Creativity*, *9*, 46–58.

Libby, R., & Libby, P. A. (1989). Expert measurement and mechanical combination in control reliance decisions. *The Accounting Review*, *64*(4): 729-747.

Low, K. Y. (2004). The Effects of Industry Specialization on Audit Risk Assessments and Audit-Planning Decisions. *The Accounting Review*, *79*(1), 201–219.

Low, K.-Y., & Tan, H.-T. (2011). Does Time Constraint Lead to Poorer Audit Performance? Effects of Forewarning of Impending Time Constraints and Instructions. *Auditing: A Journal of Practice & Theory*, *30*(4), 173–190.

Lynch, A. L., Murthy, U. S., & Engle, T. J. (2009). Fraud brainstorming using computer-mediated communication: The effects of brainstorming technique and facilitation. *The Accounting Review*, *84*(4), 1209-1232.

McCracken, S., Salterio, S. E., & Gibbins, M. (2008). Auditor–client management relationships and roles in negotiating financial reporting. *Accounting, Organizations and Society*, *33*, 362–383.

Messier, W. F., Jr, Owhoso, V., & Rakovski, C. (2008). Can Audit Partners Predict Subordinates' Ability to Detect Errors? *Journal of Accounting Research, 46*(5): 1241-1264.

Ng, T. B.-P., & Tan, H.-T. (2003). Effects of authoritative guidance availability and audit committee effectiveness on auditors' judgments in an auditor-client negotiation context. *The Accounting Review*, 801–818.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186-213.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology*, *38*(6), 535-544.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. (2003). Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, *39*(6), 531-548.

Peecher, M. E., Solomon, I., & Trotman, K. T. (2013). An accountability framework for financial statement auditors and related research questions. *Accounting, Organizations and Society*, *38*(8), 596-620.

Plumlee, D., Rixom, B. A., & Rosman, A. J. (2014) Training Auditors to Perform Analytical Procedures Using Metacognitive Skills. *The Accounting Review*, *90*(1), 351-369.

Public Company Accounting Oversight Board (PCAOB). (2003). *Using the Work of a Specialist.* PCAOB Interim Auditing Standards AU Section 336. Washington, D.C.: PCAOB.

Public Company Accounting Oversight Board (PCAOB). (2007). *Auditing Standard No. 5: An Audit of Internal Control Over Financial Reporting that is Integrated with an Audit of Financial Statements.* Washington, D.C.: PCAOB.

Public Companies Accounting Oversight Board (PCAOB). (2008). *Report on the PCAOB's 2004, 2005, 2006, and 2007 Inspections of Domestic Annually Inspected Firms*. PCAOB Release No. 2008-08. Washington, D.C.: PCAOB.

Public Company Accounting Oversight Board (PCAOB). (2010). *Auditing Standards Related to the Auditor's Assessment of and Response to Risk.* Washington, D.C.: PCAOB.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*.

Reiter-Palmon, R., & A. E. Herman (2008). Creativity and cognitive processes: Multi-level linkages between individual and team cognition. In M. D. Mumford, S. T. Hunter, & K. E. Bedell-Avers. *Research in Multi-level Linkages Between Individual and Team Cognition.* Elsevier: Oxford.

Rostan, S. M. (1994). Problem finding, problem solving, and cognitive controls: An empirical investigation of critically acclaimed productivity. *Creativity Research Journal, 7*(2), 97–110.

Simnett, R. (1996). The effect of information selection, information processing and task complexity on predictive accuracy of auditors. *Accounting, Organizations and Society*, *21*(7-8), 699–719.

Solomon, I., Shields, M. D., & Whittington, O. R. (1999). What do industry-specialist auditors know? *Journal of Accounting Research*, *37*(1), 191–208.

Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Chapter Four-Beyond Productivity Loss in Brainstorming Groups: The Evolution of a Question. *Advances in experimental social psychology*, *43*, 157-203.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463.

Trotman, K. T., Simnett, R., & Khalifa, A. (2009). Impact of the Type of Audit Team Discussions on Auditors' Generation of Material Frauds. *Contemporary Accounting Research*, *26*(4), 1115–1142.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Van Boven, L., Kane, J., & McGraw, A. P. (2009). Temporally Asymmetric Constraints on Mental Simulation: Retrospection Is More Constrained Than Prospection. In K. D. Markman, W. M. Klein, & J. A. Suhr, *Handbook of Imagination and Mental Simulation,* 131–150.

Vera-Muñoz, S. C., Kinney, W. R., Jr, & Bonner, S. E. (2001). The effects of domain experience and task presentation format on accountants' information relevance assurance. The Accounting Review, 405–429.

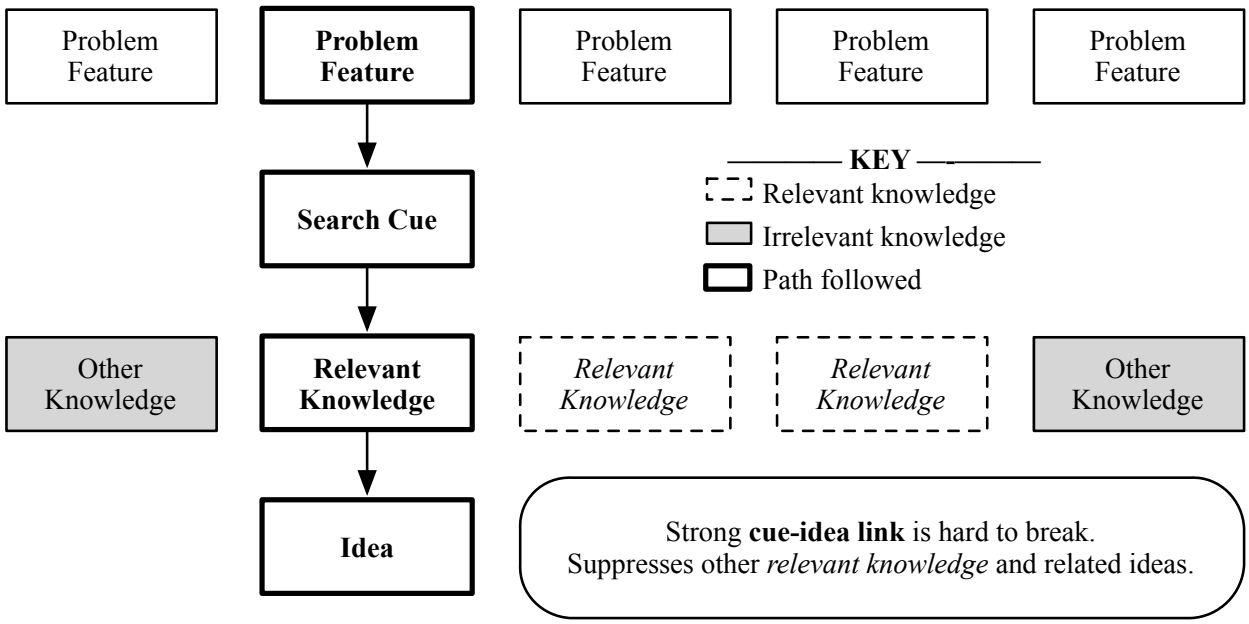
Wong, E. M., Kray, L. J., Galinsky, A. D., & Markman, K. D. (2009). Stimulating Creativity in Groups Through Mental Simulation. *Research on Managing Groups and Teams*, *12*, 111–134.

Wright, A. (1988). The impact of prior working papers on auditor evidential planning judgments. *Accounting, Organizations and Society*, *13*(6), 595–605.

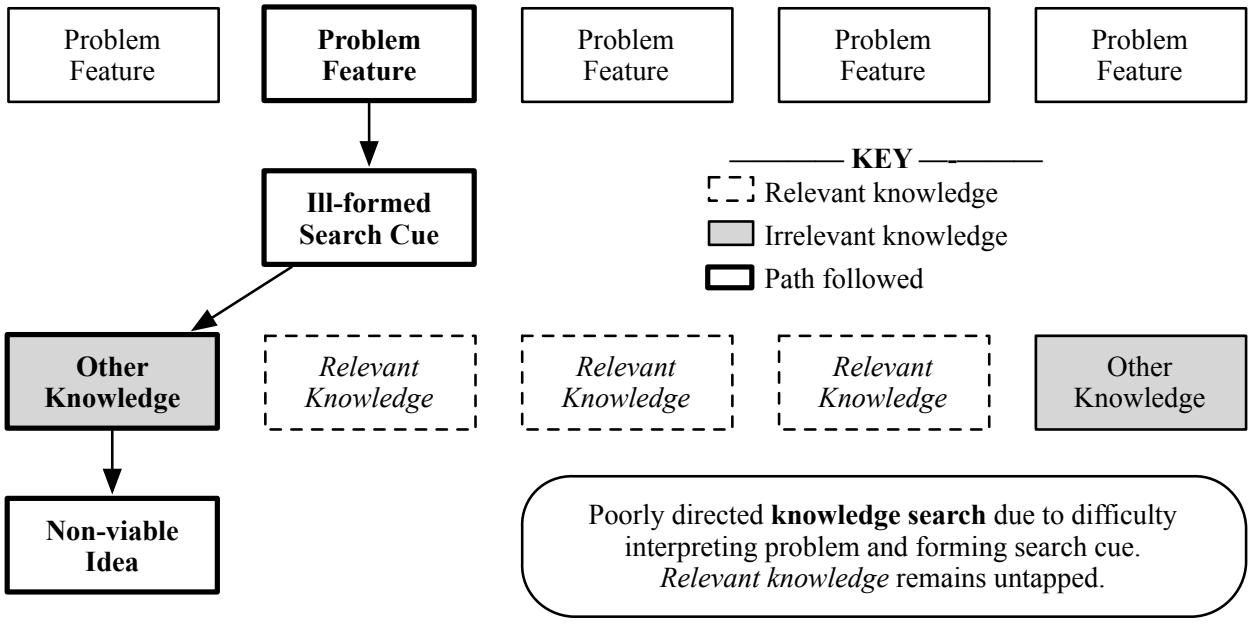
FIGURES

FIGURE 1
Search for Ideas in Associative Memory*

Panel A: Familiar Task



Panel B: Unfamiliar Task



*Adapted from Nijstad and Stroebe (2006).

FIGURE 2
Summary of Predicted Effects

Panel A: Interaction of Counterfactual Mindset and Familiarity (H1)

Familiarity Level	Knowledge Search Problem (No Intervention)	Counterfactual Mindset (CF) (Broaden Active Knowledge)	Example Provision (Improves Search Cue)	Joint Effect (Focus on Task Embedded Cues)
Lower (Unfamiliar)	*Unfocused*	Not beneficial or detrimental	Strongly beneficial	Not beneficial
Higher (Familiar)	*Too focused*	Strongly beneficial	Weakly beneficial	Strongly beneficial

H1

Panel B: Interaction of Counterfactual Mindset and Example (H2)

Familiarity Level	Knowledge Search Problem (No Intervention)	Counterfactual Mindset (Broaden Active Knowledge)	Example Provision (Improves Search Cue)	Joint Effect (Focus on Task Embedded Cues)	
Lower (Unfamiliar)	*Unfocused*	Not beneficial or detrimental	Strongly beneficial	Not beneficial	**H2a**
Higher (Familiar)	*Too focused*	Strongly beneficial	Weakly beneficial	Strongly beneficial	**H2b**

FIGURE 3
Graphical Representation of Hypothesis 1

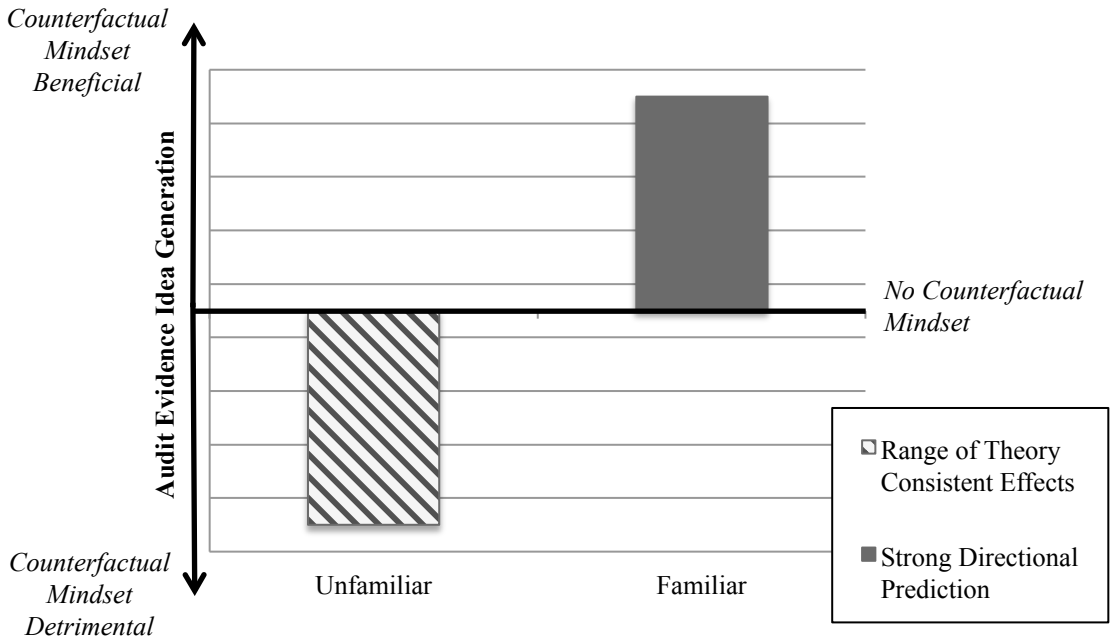
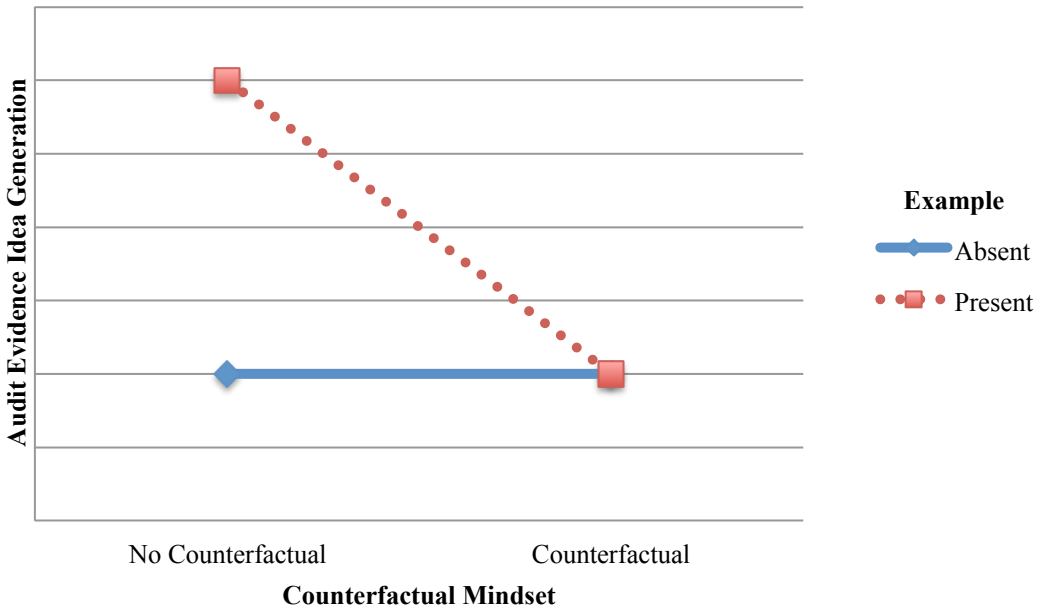


FIGURE 4
Graphical Representation of Hypothesis 2

Panel A: Lower Familiarity (H2a)



Panel B: Higher Familiarity (H2b)

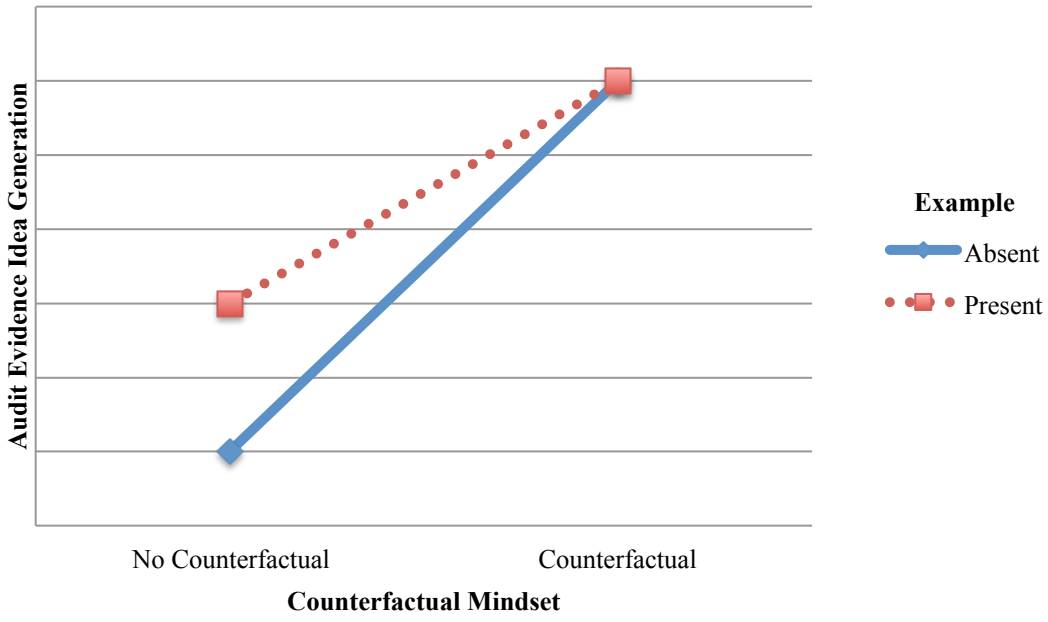


FIGURE 5
Experimental Procedures

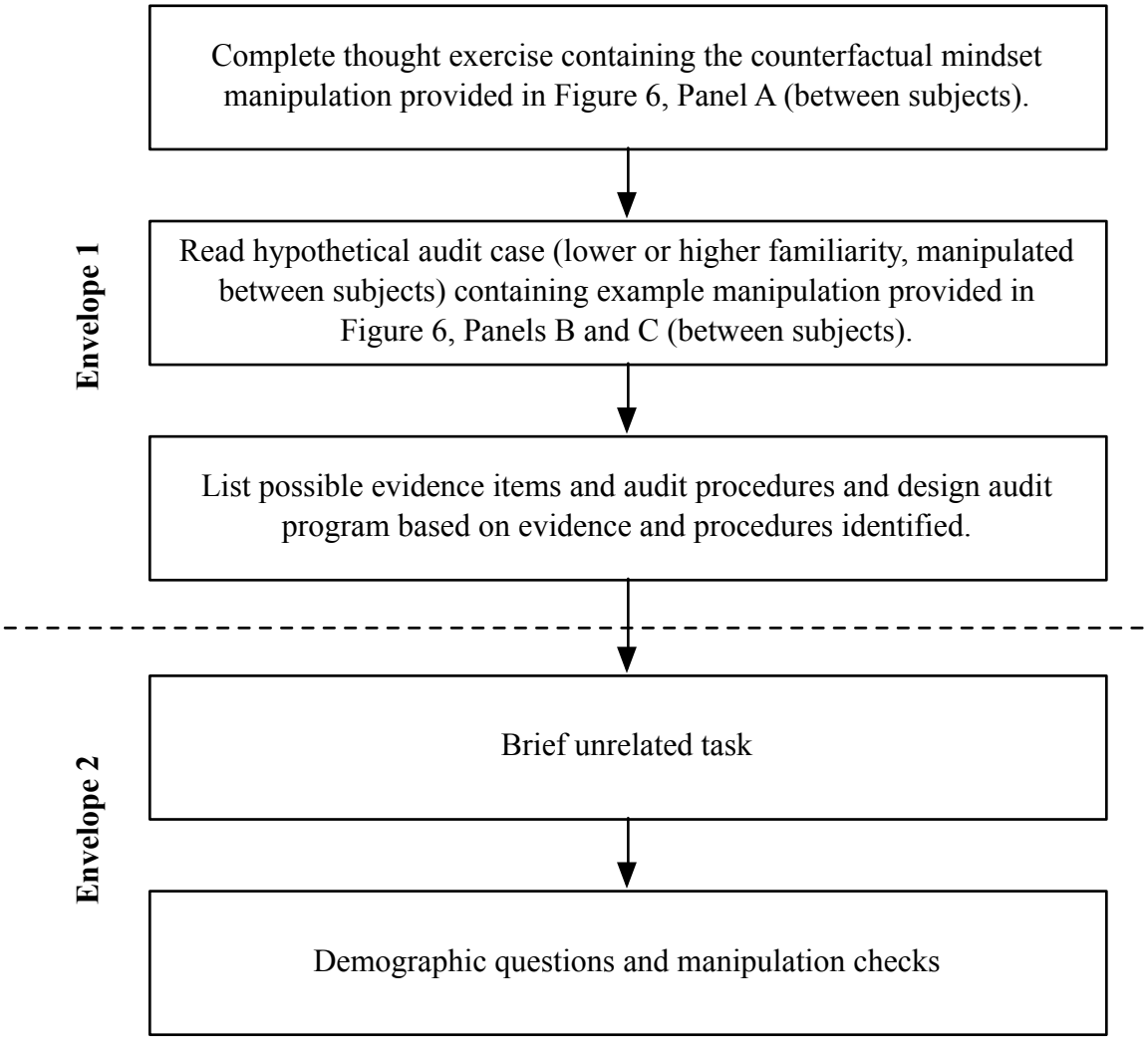


FIGURE 6
Experimental Manipulations

Panel A: Counterfactual Mindset Prime
Jane is at a rock concert of one of her favorite bands. Seating is on a first come, first served basis. At the concert the announcer reveals that a trip to Hawaii will be given to a lucky fan and that the winner will be determined by the seat number currently occupied. *[Counterfactual condition: Jane's view of the stage is partially obstructed and she sees a much better seat in the near vicinity so she changes seats.]* The announcer returns just prior to the start of the concert to announce the winner of the trip to Hawaii. To Jane's surprise the announcer calls the number of the seat she just moved to and she wins the trip to Hawaii.

Panel B: Lower Familiarity
The client is a well-known technology company that often asks job applicants very challenging interview questions. One of these questions asks the applicants, "How many golf balls would fit inside a school bus?"

The legal department of this company is concerned that disgruntled applicants could sue the company if they feel their answers are incorrectly dismissed. Therefore, the lawyers of the firm required that the company purchase a school bus and fill it with golf balls when they started using this question in interviews. Your firm has been engaged to attest as to the number of golf balls currently held inside the school bus.

Your task is to create an audit plan for this engagement by identifying audit evidence you could obtain and audit procedures that you could perform in order to get assurance over the number of golf balls in the school bus.

[Example Present Condition: For example, obtain the dimensions of the school bus and use that information in an analytical procedure to estimate the number of golf balls that the bus can hold.]

Panel C: Higher Familiarity
The client is a well-known construction company that is diversifying its operations. In the past year this company purchased a toll bridge that was previously owned and operated by a smaller private company. Toll bridge operations are not an area in which this company has experience, so these operations are completely new to the current year audit.

Your firm has been engaged to attest to the post acquisition financial statements of this company. More specifically, you have been assigned to work on the toll bridge business segment exclusively.

Your task is to create an audit plan for this engagement by identifying audit evidence you could obtain and audit procedures that you could perform in order to get assurance over the revenues associated with the toll bridge's operations.

[Example Present Condition: For example, obtain data on average daily traffic flow across the bridge and use that information in an analytical procedure to estimate the expected annual revenues.]

FIGURE 7
Example Responses for Lower Familiarity (School Bus) Case

Panel A: Maximum Factor Score (20.8) – Participant #135

-Idea 1: Get weight of school bus before & after bus is filled with golf balls. Get evidence of type of golf balls purchased (receipt, email, correspondence, etc.) and do independent test as to average weight of golf balls. Observe weight being measured & that bus was full. Based on these facts. Calculate estimate as to # of balls in bus once full.
-Idea 2: Schedule observation of these procedures & observe as the balls are counted & placed in bus.
-Idea 3: Get dimensions of bus & evidence of type of balls used. Using box, independently determine how many balls fit in box dimensions -> extrapolate to full bus dimensions.

Scoring Summary: Fluency = 10; Breadth = 4; Elaboration = 10; Originality (not part of factor) = 0.1

Panel B: Minimum Factor Score (1.7) – Participant #82

Perform a cycle count/observation count.

Scoring Summary: Fluency = 1; Breadth = 1; Elaboration = 0; Originality (not part of factor) = 0

Panel C: Maximum Originality (3.8) – Participant #117

Evidence Items
-Title to the school bus purchased by the company
-Receipts (& evidence of payment such as check copy or wire debits) from the store where the golf balls were purchased
-Picture of the school bus filled with golf balls
-Company's documentation of how many golf balls they were able to fit in the school bus
Audit Procedures
-You could reperform the task - use the same school bus & golf balls to see if your count is reasonably close to the final determination of the company (+/- 5%)
-You could observe the company complete the task & keep independent count of the golf balls
-You could perform analytical procedures: obtain dimensions of school bus & golf ball, then divide to determine approx. how many golf ball fit in the bus

Scoring Summary (Non-Originality Measures): Fluency = 10; Breadth = 3; Elaboration = 6

Panel D: Minimum Originality (0.0) – Participant #16

Evidence:
-Dimensions of golf balls/bus
Procedures:
-Weigh each golf ball to ensure each is the same size

Scoring Summary (Non-Originality Measures): Fluency = 2; Breadth = 2; Elaboration = 3

FIGURE 8
Example Responses for Higher Familiarity (Toll Booth) Case

Panel A: Maximum Factor Score (23.5) – Participant #119

1. Prior year financial statements (including BS, IS, CF)
2. Evidence obtained via discussion w/ toll bridge operator, financial manager of toll bridge
3. Support for controls in place over AR & revenue
-AR Reconciliation
-Daily sales review reports (25 samples)
-Systems support showing automatic journal entry of revenue and AR (1 sample)
-Systems support showing how revenue is recorded on toll booth (i.e., is each entry recorded or is there a batch entry/day) (1 sample)
-Count of how many cars pass each day vs. revenue recognized (25 samples)
4. Support for testing of revenue as of YE
-AR Rec @ December
-AR subledger
-5 days prior to YE and 5 days after listing, 5 selections of cars prior to YE - receipt of $, 5 selections of cars after YE - receipt of $

Scoring Summary: Fluency =11; Breadth = 3; Elaboration = 13; Originality (not part of factor) = 2.5

Panel B: Minimum Factor Score (.934) – Participant #15

1) Identify types of revenues
2) Walkthrough to identify any risks/considerations of fraud within significant types of revenues
3) Design an audit program to perform substantive testing over those revenues that properly address the risk of material misstatements identified in #2 above
4) Test substantively

Scoring Summary: Fluency = 1; Breadth = 0; Elaboration = 0; Originality (not part of factor) = 0.0

Panel C: Maximum Originality (8.5) – Participant #40

-BoD minutes
-Revenue projections
-Consultant reports
-Confirmation with/from vendors

Scoring Summary (Non-Originality Measures): Fluency = 2; Breadth = 1; Elaboration = 0

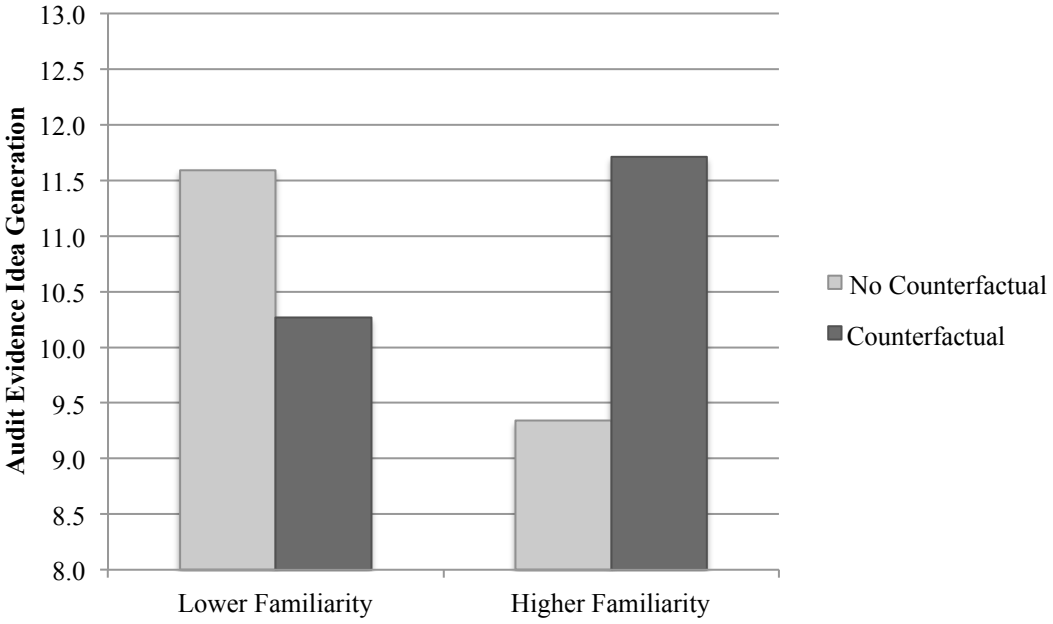
Panel D: Minimum Originality (0.0) – Participant #22

-Test of details on cash
-Analyze somehow traffic flow per hour. Then multiply by price.

Scoring Summary (Non-Originality Measures): Fluency = 1; Breadth = 1; Elaboration = 1

FIGURE 9
Graphical Representation of Results – Idea Generation[a]

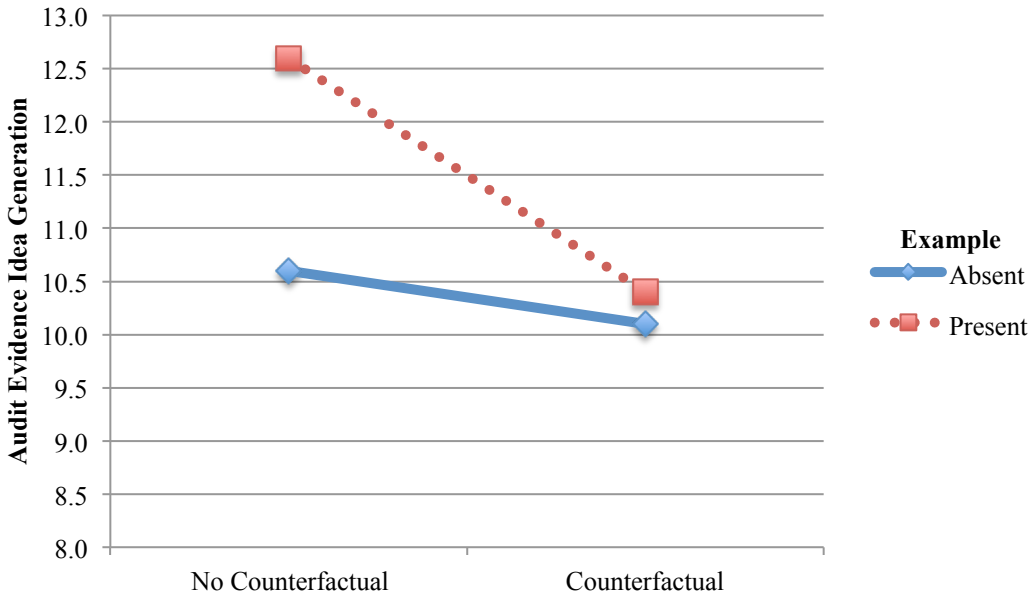
Panel A: Interaction of Counterfactual Mindset and Familiarity (H1)



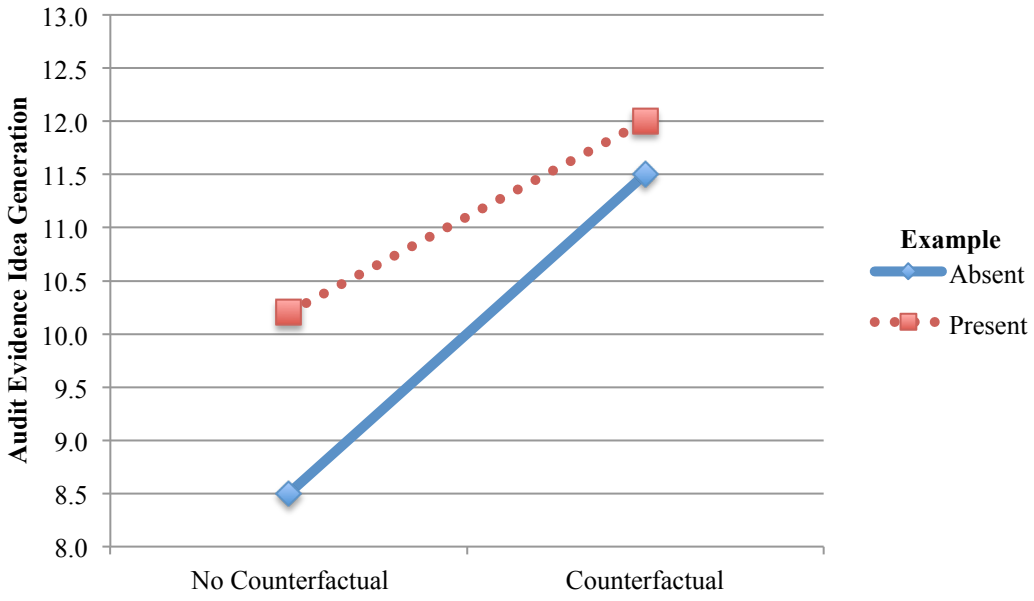
Notes:
[a] I measure audit evidence idea generation as a composite measure of the sum of fluency (number of unique ideas), breadth (variety of ideas) and elaboration (level of idea specification) of each idea as coded independently by the author and a second coder and weighted by factor analysis loadings.

FIGURE 10
Graphical Representation of H2 Results by Familiarity Condition– Idea Generation[a]

Panel A: Lower Familiarity (H2a)



Panel B: Higher Familiarity (H2b)



Notes:
[a] I measure audit evidence idea generation as a composite measure of the sum of fluency (number of unique ideas), breadth (variety of ideas) and elaboration (level of idea specification) of each idea as coded independently by the author and a second coder and weighted by factor analysis loadings.

TABLES

TABLE 1
Interview Participant Demographic Information

	All	Seniors	Managers	Partners
	n=15	n=5	n=5	n=5
		Mean (Std. Dev.)		
Age	35.7 (9.3)	30.0 (7.0)	31.4 (2.7)	45.8 (7.6)
Years as an Auditor	11.4 (9.8)	3.4 (0.6)	8.4 (3.2)	22.4 (9.0)
		Frequency (Percent of n)		
Firm size				
Big 4	12 (80%)	3 (60%)	5 (100%)	4 (80%)
Midsize Regional	3 (20%)	2 (40%)	0 (0%)	1 (20%)
Industry of specialization				
Financial Services	6 (40%)	2 (40%)	1 (20%)	3 (60%)
Insurance	3 (20%)	2 (40%)	1 (20%)	0 (0%)
Real Estate	3 (20%)	0 (0%)	2 (40%)	1 (20%)
Manufacturing	2 (13%)	1 (20%)	1 (20%)	0 (0%)
Public Sector	1 (7%)	0 (0%)	0 (0%)	1 (20%)

TABLE 2
Interview Examples

	Easy Task	Difficult Task
Senior 1	Confirming investments	Auditing allowance for loan losses
Senior 2	Auditing cash	Auditing a loss share agreement with FDIC
Senior 3	Testing fixed asset additions	Confirming reinsurance receivables
Senior 4	Testing cash reconciliations	Auditing insurance loss adjustment reserves
Senior 5	Inventory test counts	Reviewing diluted EPS calculation
Manager 1	Testing income journal activity (validating third party work)	Auditing securities lending transactions
Manager 2	Testing fixed asset additions	Inventory price testing
Manager 3	Auditing cash	Goodwill impairment testing
Manager 4	Reviewing discontinued operations disclosure presentation	Lease contract impairment testing
Manager 5	Auditing revolving line of credit	Auditing litigation reserves related to natural disaster
Partner 1	Testing information technology controls	Testing information technology access for nonstandard process
Partner 2	Reviewing Pro-forma disclosures for IPO	Reviewing purchase price allocations for IPO
Partner 3	Vouching capital activity	Valuation of Level 3 securities
Partner 4	Testing mortgage servicing rights (validating third party work)	Auditing allowance for loan losses
Partner 5	Confirming cash	Evaluating accounting treatment of contribution to joint venture

TABLE 3
Factor Analysis of Survey Responses

Panel A: Factor Analysis for Difficult Scenarios – Quartimax Rotation

#	Question	Factor 1	Factor 2	Factor 3	Factor 4
Task-Internal Factor					
4	How much professional judgment was required to execute the audit program procedures surrounding this transaction?	0.98			
1	How routine was the transaction for the client being audited?	0.92			
5	On average, how subjective or objective was the evidence you were reviewing?	0.67			
6	For any evidence prepared by other members of your firm, on average what was their level of expertise for this type of transaction?	0.67			
Environmental Factor					
9	To what extent did management express a preferred accounting treatment that either agreed or disagreed with the initial position of your firm?		0.87		
10	How similar was the final working paper to the prior year's working paper?		0.75		
11	What was the status of the engagement's budget at the time this issue was reviewed?		0.75		
Task-External Factor					
3	To what degree was the applicable accounting guidance unambiguous in that it clearly indicated one correct accounting treatment as opposed to allowing multiple possible correct methods of accounting?			0.98	
Person Factor					
2	How much personal experience did you have with auditing transactions like this?				0.96
	Eigenvalue	2.99	2.31	1.29	1.00
	Percentage variance explained	30.77%	25.64%	15.33%	12.60%

TABLE 3 (cont.)
Factor Analysis of Survey Responses

Panel B: Descriptive Statistics Regarding Task-Internal Factor - Mean (Std. Dev.) [n]

	Rank											
	Senior				Manager				Partner			
Difficulty	Q1	Q4	Q5	Q6	Q1	Q4	Q5	Q6	Q1	Q4	Q5	Q6
Easy	0.3	3.6	0.6	1.5	1.4	2.1	1.4	4.4	4.1	4.6	3.0	4.6
	(0.6)	(0.8)	(0.5)	(0.8)	(0.6)	(0.8)	(0.5)	(0.8)	(0.6)	(0.8)	(0.5)	(0.8)
	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]
Difficult	5.9	7.9	3.7	5.6	7.8	8.4	5.8	6.5	4.5	8.4	5.3	7.0
	(2.0)	(0.4)	(1.6)	(1.2)	(2.0)	(0.4)	(1.6)	(1.2)	(2.0)	(0.4)	(1.6)	(1.2)
	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]

Panel C: Analysis of Variance Regarding Task-Internal Factor

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	11.55	1	11.55	1.51	0.251
Rank	48.80	2	24.40	3.18	0.090
Order*Rank	34.38	2	17.19	2.24	0.162
Error	69.04	9	7.67		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Question	104.90	3	34.97	8.65	< 0.001
Question*Order	6.51	3	2.17	0.54	0.661
Question*Rank	19.46	6	3.24	0.80	0.577
Question*Order*Rank	49.17	6	8.20	2.03	0.097
Error	109.17	27	4.04		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	409.81	1	409.81	23.06	0.001
Difficulty*Order	0.12	1	0.12	0.01	0.936
Difficulty*Rank	35.08	2	17.54	0.99	0.410
Difficulty*Order*Rank	36.51	2	18.26	1.03	0.396
Error	159.92	9	17.77		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Question*Difficulty	17.04	3	5.68	1.71	0.189
Question*Difficulty*Order	14.83	3	4.94	1.49	0.240
Question*Difficulty*Rank	34.35	6	5.73	1.72	0.154
Question*Difficulty*Order*Rank	12.70	6	2.12	0.64	0.700
Error	89.70	27	3.32		

Notes: All p-values are two-tailed.
Dependent: Routineness [Q1], Professional Judgment [Q4], Evidence Subjectivity [Q5], Expertise [Q6]; Factors: Order, Rank, Difficulty

TABLE 3 (cont.)
Factor Analysis of Survey Responses

Panel D: Descriptive Statistics Regarding Environmental Factor - Mean (Std. Dev.) [n]

	Rank								
	Senior			Manager			Partner		
Difficulty	Q9	Q10	Q11	Q9	Q10	Q11	Q9	Q10	Q11
Easy	3.6	2.3	5.0	3.5	3.0	5.0	3.8	1.0	5.0
	(1.5)	(1.4)	(0.1)	(1.9)	(1.7)	(0.1)	(2.0)	(1.8)	(0.1)
	[5]	[5]	[5]	[4]	[4]	[4]	[3]	[3]	[3]
Difficult	5.2	5.3	5.8	4.7	5.0	5.0	8.3	6.0	7.8
	(0.7)	(1.2)	(0.5)	(0.9)	(1.5)	(0.6)	(1.0)	(1.6)	(0.7)
	[5]	[5]	[5]	[4]	[4]	[4]	[3]	[3]	[3]

Panel E: Analysis of Variance Regarding Environmental Factor

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	0.11	1	0.11	0.02	0.907
Rank	8.78	2	4.39	0.59	0.585
Order*Rank	2.17	2	1.08	0.15	0.868
Error	44.81	6	7.47		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Question	31.59	2	15.79	3.72	0.055
Question*Order	2.06	2	1.03	0.24	0.788
Question*Rank	10.31	4	2.58	0.61	0.665
Question*Order*Rank	20.19	4	5.05	0.36	0.364
Error	50.89	12	4.24		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	77.78	1	77.78	12.04	0.013
Difficulty*Order	0.54	1	0.54	0.08	0.783
Difficulty*Rank	21.22	2	10.61	1.64	0.270
Difficulty*Order*Rank	29.49	2	14.74	2.28	0.183
Error	38.76	6	6.46		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Question*Difficulty	11.57	2	5.78	1.56	0.251
Question*Difficulty*Order	1.80	2	0.90	0.24	0.788
Question*Difficulty*Rank	0.57	4	0.14	0.04	0.997
Question*Difficulty*Order*Rank	16.90	4	4.23	1.14	0.386
Error	44.63	12	3.72		

Notes: All p-values are two-tailed.
Dependent: Preference [Q9], Similarity [Q10], Budget Status [Q11]; Factors: Order, Rank, Difficulty)

TABLE 4

Survey Question: How much professional judgment was required to execute the audit program procedures surrounding this transaction?

(Dependent: Professional Judgment; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	3.6	2.1	4.6	3.5
	(0.8)	(0.8)	(0.8)	
	n=5	n=5	n=5	n=15
Difficult	7.9	8.4	8.4	8.2
	(0.4)	(0.4)	(0.4)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	0.00	1	0.00	0.00	0.989
Rank	7.11	2	3.56	3.11	0.094
Order*Rank	15.73	2	7.86	6.87	0.015
Error	10.30	9	1.15		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	167.81	1	167.81	64.74	0.000
Difficulty*Order	0.03	1	0.03	0.01	0.921
Difficulty*Rank	7.89	2	3.94	1.52	0.270
Difficulty*Order*Rank	16.64	2	8.32	3.21	0.089
Error	23.33	9	2.59		

Note: All p-values are two-tailed.

TABLE 5
Survey Question: How routine was this transaction for the client being audited?
(Dependent: Routineness; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	0.3	1.4	4.1	1.8
	(0.6)	(0.6)	(0.6)	
	n=5	n=5	n=5	n=15
Difficult	5.9	7.8	4.5	5.9
	(2.0)	(2.0)	(2.0)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	6.73	1	6.73	0.70	0.425
Rank	12.05	2	6.02	0.63	0.556
Order*Rank	23.24	2	11.62	1.21	0.343
Error	86.55	9	9.61		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	124.00	1	124.00	11.56	0.008
Difficulty*Order	9.89	1	9.89	0.92	0.362
Difficulty*Rank	50.74	2	25.37	2.37	0.149
Difficulty*Order*Rank	17.89	2	8.94	8.94	0.465
Error	96.54	9	10.73	10.73	

Note: All p-values are two-tailed.

TABLE 6
Survey Question: On average, how subjective or objective was the evidence that you were reviewing?
(Dependent: Evidence Subjectivity; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	0.6	1.4	3.0	1.6
	(0.5)	(0.5)	(0.5)	
	n=5	n=5	n=5	n=15
Difficult	3.7	5.8	5.3	4.8
	(1.6)	(1.6)	(1.6)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	1.01	1	1.01	0.18	0.684
Rank	21.49	2	10.75	1.88	0.208
Order*Rank	5.49	2	2.74	0.48	0.634
Error	51.54	9	5.73		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	76.44	1	76.44	9.34	0.014
Difficulty*Order	0.00	1	0.00	0.00	0.990
Difficulty*Rank	5.28	2	0.32	0.32	0.732
Difficulty*Order*Rank	9.24	2	0.57	0.57	0.587
Error		9			

Note: All p-values are two-tailed.

TABLE 7

Survey Question: For any evidence prepared by other members of your firm, on average what was their level of expertise for this type of transaction?

(Dependent: Expertise; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	1.5	4.4	4.6	3.5
	(0.8)	(0.8)	(0.8)	
	n=5	n=5	n=5	n=15
Difficult	5.6	6.5	7.0	6.5
	(1.2)	(1.2)	(1.2)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	10.32	1	10.32	3.11	0.111
Rank	27.61	2	13.80	4.17	0.052
Order*Rank	39.10	2	19.55	5.90	0.023
Error	29.82	9	3.31		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	58.60	1	58.60	9.40	0.013
Difficulty*Order	5.03	1	5.03	0.81	0.392
Difficulty*Rank	5.53	2	2.77	0.44	0.655
Difficulty*Order*Rank	5.44	2	2.72	0.44	0.659
Error	56.12	9	6.24		

Note: All p-values are two-tailed.

TABLE 8

Survey Question: To what degree was the applicable accounting guidance unambiguous in that it clearly indicated one correct accounting treatment as opposed to allowing multiple possibly correct methods of accounting?

(Dependent: Guidance Ambiguity; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	2.4	1.2	3.5	2.5
	(1.1)	(1.1)	(1.1)	
	n=5	n=5	n=5	n=15
Difficult	3.5	5.4	6.4	5.4
	(1.0)	(1.0)	(1.0)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	1.70	1	1.70	0.46	0.514
Rank	22.35	2	11.18	3.03	0.099
Order*Rank	40.63	2	20.32	5.50	0.028
Error	33.24	9	3.69		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	54.12	1	54.12	7.52	0.023
Difficulty*Order	0.24	1	0.24	0.03	0.861
Difficulty*Rank	11.63	2	5.82	0.81	0.475
Difficulty*Order*Rank	9.77	2	4.88	0.68	0.531
Error	64.74	9	7.19		

Note: All p-values are two-tailed.

TABLE 9

Survey Question: To what extent did management express a preferred accounting treatment that either agreed or disagreed with the initial position of your firm?

(Dependent: Preference; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	3.6	3.5	3.8	3.5
	(1.4)	(1.8)	(1.5)	
	n=5	n=4	n=4	n=13
Difficult	5.2	4.7	7.8	5.6
	(0.7)	(0.9)	(0.8)	
	n=5	n=4	n=4	n=13

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	1.58	1	1.58	0.26	0.623
Rank	12.01	2	6.01	1.00	0.415
Order*Rank	6.25	2	3.12	0.52	0.615
Error	41.99	7	6.00		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	28.78	1	28.78	5.00	0.061
Difficulty*Order	0.74	1	0.74	0.13	0.730
Difficulty*Rank	8.94	2	4.47	0.78	0.496
Difficulty*Order*Rank	39.18	2	19.59	3.40	0.093
Error	40.33	7	5.76		

Note: All p-values are two-tailed.

TABLE 10
Survey Question: How similar was the final working paper to the prior year's working paper?
(Dependent: Similarity; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	2.3	3.0	1.5	2.7
	(1.2)	(1.2)	(1.6)	
	n=5	n=5	n=4	n=14
Difficult	5.3	6.5	5.8	5.9
	(1.2)	(1.2)	(1.5)	
	n=5	n=5	n=4	n=14

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	0.17	1	0.17	0.02	0.886
Rank	5.87	2	2.93	0.38	0.694
Order*Rank	9.32	2	4.66	0.61	0.568
Error	61.41	8	7.68		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	77.76	1	77.76	11.35	0.010
Difficulty*Order	0.02	1	0.02	0.00	0.960
Difficulty*Rank	1.60	2	0.80	0.12	0.891
Difficulty*Order*Rank	36.88	2	18.44	2.69	0.128
Error	54.81	8	6.85		

Note: All p-values are two-tailed.

TABLE 11

Survey Question: What was the status of the engagement's budget at the time when this issue was reviewed?

(Dependent: Budget Status; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	5.0	5.0	5.0	5.0
	(< 0.0)	(< 0.0)	(< 0.0)	
	n=5	n=5	n=5	n=15
Difficult	5.8	5.7	7.6	6.4
	(0.5)	(0.5)	(0.5)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	2.54	1	2.54	3.67	0.088
Rank	5.83	2	2.92	4.21	0.051
Order*Rank	0.31	2	0.16	0.23	0.802
Error	6.23	9	0.69		

Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	13.67	1	13.67	21.42	0.001
Difficulty*Order	2.13	1	2.13	3.34	0.101
Difficulty*Rank	5.69	2	2.84	4.46	0.045
Difficulty*Order*Rank	0.43	2	0.21	0.34	0.723
Error	5.74	9	0.64		

Note: All p-values are two-tailed.

TABLE 12
Survey Question: How much personal experience did you have with auditing transactions like this?
(Dependent: Experience; Factors: Order, Rank, Difficulty)

Panel A: Descriptive Statistics - Mean (Std. Dev.)

	Rank			
Difficulty	**Senior**	**Manager**	**Partner**	**Total**
Easy	9.0	8.7	8.8	8.7
	(0.7)	(0.7)	(0.7)	
	n=5	n=5	n=5	n=15
Difficult	4.5	5.1	7.8	5.9
	(1.5)	(1.5)	(1.5)	
	n=5	n=5	n=5	n=15

Panel B: Analysis of Variance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Order	0.05	1	0.05	0.01	0.940
Rank	14.42	2	7.21	0.87	0.450
Order*Rank	1.55	2	0.77	0.09	0.911
Error	74.19	9	8.24		
Source (within subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Difficulty	64.80	1	64.80	12.76	0.006
Difficulty*Order	1.42	1	1.42	0.28	0.609
Difficulty*Rank	16.00	2	8.00	1.58	0.259
Difficulty*Order*Rank	14.58	2	7.29	1.44	0.288
Error	45.70	9	5.08		

Note: All p-values are two-tailed.

TABLE 13
Idea Generation Descriptive Statistics[a]

Audit Evidence Idea Generation (Standard Error) by Task Condition

	Lower Familiarity		
	Counterfactual Mindset		
Example	**No Counterfactual**	**Counterfactual**	**Combined**
Absent	10.60	10.11	10.35
	(1.08)	(1.11)	(0.77)
	n=18	n=17	n=35
Present	12.59	10.44	11.51
	(1.12)	(1.11)	(0.79)
	n=17	n=17	n=34
Combined	11.59	10.27	
	(0.78)	(0.78)	
	n=35	n=34	

	Higher Familiarity		
	Counterfactual Mindset		
Example	**No Counterfactual**	**Counterfactual**	**Combined**
Absent	8.50	11.46	9.98
	(1.08)	(1.11)	(0.77)
	n=18	n=17	n=35
Present	10.17	11.97	11.07
	(1.11)	(1.11)	(0.79)
	n=17	n=17	n=34
Combined	9.34	11.71	
	(0.78)	(0.78)	
	n=35	n=34	

Notes:
[a] I measure audit evidence idea generation as a composite measure of the sum of fluency (number of unique ideas), breadth (variety of ideas) and elaboration (level of idea specification) of each idea as coded independently by the author and a second coder and weighted by factor analysis loadings.

TABLE 14
Inferential Statistics Regarding Overall Audit Evidence Idea Generation[a]
(Dependent: Audit Evidence Idea Generation; Factors: Example, Counterfactual Mindset, Task; Covariate: Time)

Panel A: Analysis of Covariance

Source (between subjects)	SS	df	MS	F-Statistic	p-value
Task[b]	5.67	1	5.67	0.27	0.603
Counterfactual Mindset[c]	9.63	1	9.63	0.46	0.498
Example[d]	43.82	1	43.82	2.10	0.150
Task by Counterfactual Mindset **(H1)**	115.99	1	115.99	5.56	**0.010**
Task by Example	0.04	1	0.04	< 0.01	0.965
Counterfactual Mindset by Example	17.08	1	17.08	0.82	0.367
Task by Counterfactual Mindset by Example	0.56	1	0.56	0.03	0.870
Time[e]	526.92	1	526.92	25.24	< 0.001
Error	2,692.88	129	20.88		

Panel B: Follow-up Simple Effects for H1

Simple Main Effects of Counterfactual by Task	SS	df	MS	F-Statistic	p-value
Audit evidence idea generation is greater for counterfactual versus no counterfactual given higher familiarity	97.30	1	97.30	4.66	**0.017**
Audit evidence idea generation is greater for counterfactual versus no counterfactual given lower familiarity	29.53	1	29.53	1.42	0.236

Panel C: Planned Contrast Interaction of Counterfactual and Example

Lower Familiarity	SS	df	MS	F-Statistic	p-value
H2a: When familiarity is lower the provision of an example will result in the greatest idea generation (+3) versus providing no intervention (-1), a counterfactual mindset prompt in isolation (-1), or both prompts together (-1).	1,063.75	1	1,063.75	50.96	**< 0.001**

Higher Familiarity	SS	df	MS	F-Statistic	p-value
H2b: When familiarity is higher the provision of an example (-1) will result in greater idea generation versus providing no intervention (-3) and prompting a counterfactual mindset will result in the greatest idea generation regardless of example presence (+2) or absence (+2).	460.96	1	460.96	22.08	**< 0.001**

Notes:
Reported p-values are two-tailed unless testing a one-tailed prediction, as signified in bold face.
[a] I measure audit evidence idea generation as a composite measure of the sum of fluency (number of unique ideas), breadth (variety of ideas) and elaboration (level of idea specification) of each idea as coded independently by the author and a second coder and weighted by factor analysis loadings.
[b] Task = case contained *lower familiarity* (school bus) versus *higher familiarity* (toll booth), manipulated between subjects
[c] Counterfactual Mindset = *counterfactual* mindset prime versus *no counterfactual* mindset prime, manipulated between subjects
[d] Example = example audit evidence was *present* verses *absent*, manipulated between subjects
[e] Time = amount of time in minutes that the participant used to complete the experiment in its entirety

TABLE 15
Originality Descriptive Statistics[a]

Originality (Standard Error) by Task Condition

	Lower Familiarity		
	Counterfactual Mindset		
Example	**No Counterfactual**	**Counterfactual**	**Combined**
Absent	1.41	1.36	1.38
	(0.40)	(0.40)	(0.28)
	n=18	n=17	n=35
Present	1.09	1.27	1.18
	(0.40)	(0.40)	(0.28)
	n=17	n=17	n=34
Combined	1.25	1.31	
	(0.28)	(0.28)	
	n=35	n=34	

	Higher Familiarity		
	Counterfactual Mindset		
Example	**No Counterfactual**	**Counterfactual**	**Combined**
Absent	2.53	2.19	2.36
	(0.39)	(0.40)	(0.28)
	n=18	n=17	n=35
Present	2.21	1.58	1.89
	(0.40)	(0.40)	(0.28)
	n=17	n=17	n=34
Combined	2.37	1.89	
	(0.28)	(0.28)	
	n=35	n=34	

Notes:
[a] I measure the originality, or the novelty of proposed audit evidence items generated, as the sum of the number of points awarded to each response for rarity in comparison to the entire pool of participant responses divided by the total number of ideas (fluency).

TABLE 16
Inferential Statistics Regarding Audit Evidence Originality[a]
(Dependent: Originality; Factors: Example, Counterfactual Mindset, Task;
Covariate: Time)

Panel A: Analysis of Covariance

Source (between subjects)	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Task[b]	24.47	1	24.47	9.17	0.003
Counterfactual Mindset[c]	1.45	1	1.45	0.54	0.463
Example[d]	3.87	1	3.87	1.45	0.231
Task by Counterfactual Mindset **(H1)**	2.54	1	2.54	0.95	0.331
Task by Example	0.58	1	0.58	0.22	0.642
Counterfactual Mindset by Example	0.01	1	0.01	< 0.01	0.956
Task by Counterfactual Mindset by Example	0.59	1	0.59	0.22	0.640
Time	3.00	1	3.00	1.12	0.291
Error	344.36	129	2.67		

Panel B: Follow-up Simple Effects for H1

Simple Main Effects of Counterfactual by Task	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
Originality is greater for counterfactual versus no counterfactual given higher familiarity	3.95	1	3.95	1.48	0.226
Originality is greater for counterfactual versus no counterfactual given lower familiarity	0.08	1	0.08	0.03	0.863

Panel C: Planned Contrast Interaction of Counterfactual and Example

Lower Familiarity	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
H2a: When familiarity is lower the provision of an example will result in the greatest originality (+3) versus providing no intervention (-1), a counterfactual mindset prompt in isolation (-1), or both prompts together (-1).	9.31	1	9.31	3.49	0.064

Higher Familiarity	**SS**	**df**	**MS**	**F-Statistic**	**p-value**
H2b: When familiarity is higher the provision of an example (-1) will result in greater originality versus providing no intervention (-3) and prompting a counterfactual mindset will result in the greatest originality regardless of example presence (+2) or absence (+2).	0.27	1	0.27	0.10	0.749

Notes:
Reported p-values are two-tailed unless testing a one-tailed prediction, as signified in bold face.
[a] I measure the originality, or the novelty of proposed audit evidence items generated, as the sum of the number of points awarded to each response for rarity in comparison to the entire pool of participant responses divided by the total number of ideas (fluency).
[b] Task = case contained *lower familiarity* (school bus) versus *higher familiarity* (toll booth), manipulated between subjects
[c] Counterfactual Mindset = *counterfactual* mindset prime versus *no counterfactual* mindset prime, manipulated between subjects
[d] Example = example audit evidence was *present* verses *absent*, manipulated between subjects
e Time = amount of time in minutes that the participant used to complete the experiment in its entirety

TABLE 17
Supplemental Descriptive Statistics of Idea Generation Component Measures

Panel A: Fluency of Proposed Audit Evidence (Standard Error)[a]

	Lower Familiarity		Higher Familiarity	
	Counterfactual Mindset		Counterfactual Mindset	
Example	No Counterfactual	Counterfactual	No Counterfactual	Counterfactual
Absent	5.0	4.8	4.8	5.6
	(0.5)	(0.6)	(0.5)	(0.6)
	n=18	n=17	n=18	n=17
Present	5.8	4.8	4.8	5.8
	(0.6)	(0.6)	(0.6)	(0.6)
	n=17	n=17	n=17	n=17

Panel B: Breadth of Proposed Audit Evidence (Standard Error)[b]

	Lower Familiarity		Higher Familiarity	
	Counterfactual Mindset		Counterfactual Mindset	
Example	No Counterfactual	Counterfactual	No Counterfactual	Counterfactual
Absent	2.8	2.8	2.3	2.3
	(0.3)	(0.3)	(0.3)	(0.3)
	n=18	n=17	n=18	n=17
Present	3.5	3.1	2.2	2.6
	(0.3)	(0.3)	(0.3)	(0.3)
	n=17	n=17	n=17	n=17

Panel C: Elaboration of Proposed Audit Evidence (Standard Error)[c]

	Lower Familiarity		Higher Familiarity	
	Counterfactual Mindset		Counterfactual Mindset	
Example	No Counterfactual	Counterfactual	No Counterfactual	Counterfactual
Absent	4.6	4.2	2.8	5.3
	(0.6)	(0.7)	(0.6)	(0.7)
	n=18	n=17	n=18	n=17
Present	5.4	4.3	4.8	5.5
	(0.6)	(0.7)	(0.7)	(0.7)
	n=17	n=17	n=17	n=17

Notes:
[a] I measure the fluency, or the quantity of audit evidence ideas generated, as the sum of the number of evidence items generated as independently coded by the author and a second coder.
[b] I measure the breadth, or the variety of proposed audit evidence items generated, as the sum of the number of categories spanned by the proposed audit evidence items as independently coded by the author and a second coder.
[c] I measure the elaboration, or the degree of specification of the audit evidence ideas generated, as the sum of the number of points awarded as independently coded by the author and a second coder.

TABLE 18

Inferential Statistics Regarding Fluency of Idea Generation[a]

(Dependent: Fluency of Proposed Audit Evidence; Factors: Example, Counterfactual Mindset, Task; Covariate: Time)

Panel A: Analysis of Covariance

Source (between subjects)	SS	df	MS	F-Statistic	p-value
Task[b]	0.79	1	0.79	0.15	0.701
Counterfactual Mindset[c]	0.59	1	0.59	0.11	0.739
Example[d]	2.66	1	2.66	0.50	0.481
Task by Counterfactual Mindset	20.43	1	20.43	3.84	**0.026**
Task by Example	0.76	1	0.76	0.14	0.705
Counterfactual Mindset by Example	0.91	1	0.91	0.17	0.679
Task by Counterfactual Mindset by Example	2.40	1	2.40	0.45	0.502
Time	137.69	1	137.69	25.90	<0.001
Error	685.90	129	5.32		

Panel B: Simple Effects Analogous to H1

Simple Main Effects of Counterfactual by Task	SS	df	MS	F-Statistic	p-value
Fluency of idea generation is greater for counterfactual versus no counterfactual given higher familiarity	14.16	1	14.16	2.66	**0.053**
Fluency of idea generation is greater for counterfactual versus no counterfactual given lower familiarity	7.05	1	7.05	1.33	0.252

Panel C: Planned Contrast Interaction of Counterfactual and Example (Analogous to H2)

Lower Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is lower the provision of an example will result in the greatest fluency of idea generation (+3) versus providing no intervention (-1), a counterfactual mindset prompt in isolation (-1), or both prompts together (-1).	238.91	1	238.91	44.93	**<0.001**

Higher Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is higher the provision of an example (-1) will result in greater fluency of idea generation versus providing no intervention (-3) and prompting a counterfactual mindset will result in the greatest originality regardless of example presence (+2) or absence (+2).	74.57	1	74.57	14.02	**<0.001**

Notes:

Reported p-values are two-tailed unless testing a one-tailed prediction, as signified in bold face.

[a] I measure the fluency, or the quantity of audit evidence ideas generated, as the sum of the number of evidence items generated as independently coded by the author and a second coder.

[b] Task = task contained *lower familiarity* (school bus) versus *higher familiarity* (toll booth), manipulated between subjects

[c] Counterfactual Mindset = *counterfactual* mindset prime versus *no counterfactual* mindset prime, manipulated between subjects

[d] Example = example audit evidence was present verses absent, manipulated between subjects

TABLE 19

Inferential Statistics Regarding Breadth of Proposed Audit Evidence[a]

(Dependent: Breadth of Proposed Audit Evidence; Factors: Example, Counterfactual Mindset, Task; Covariate: Time)

Panel A: Analysis of Covariance

Source (between subjects)	SS	df	MS	F-Statistic	p-value
Task[b]	16.15	1	16.15	14.00	<0.001
Counterfactual Mindset[c]	<0.01	1	<0.01	<0.01	0.953
Example[d]	2.91	1	2.91	2.52	0.115
Task by Counterfactual Mindset	0.99	1	0.99	0.86	**0.179**
Task by Example	1.66	1	1.66	1.44	0.232
Counterfactual Mindset by Example	0.02	1	0.02	0.01	0.906
Task by Counterfactual Mindset by Example	1.18	1	1.18	1.02	0.314
Time	3.95	1	3.95	3.42	0.067
Error	148.72	129	1.15		

Panel B: Simple Effects Analogous to H1

Simple Main Effects of Counterfactual by Task	SS	df	MS	F-Statistic	p-value
Breadth of proposed audit evidence is greater for counterfactual versus no counterfactual given higher familiarity	0.57	1	0.57	0.49	**0.243**
Breadth of proposed audit evidence is greater for counterfactual versus no counterfactual given lower familiarity	0.43	1	0.43	0.38	0.542

Panel C: Planned Contrast Interaction of Counterfactual and Example (Analogous to H2)

Lower Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is lower the provision of an example will result in the greatest breadth (+3) versus providing no intervention (-1), a counterfactual mindset prompt in isolation (-1), or both prompts together (-1).	68.64	1	68.64	59.54	**<0.001**

Higher Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is higher the provision of an example (-1) will result in greater breadth versus providing no intervention (-3) and prompting a counterfactual mindset will result in the greatest originality regardless of example presence (+2) or absence (+2).	10.60	1	10.60	9.19	**0.002**

Notes:

Reported p-values are two-tailed unless testing a one-tailed prediction, as signified in bold face.

[a] I measure the breadth, or variety of proposed audit evidence items generated, as the sum of the number of categories spanned by the proposed audit evidence items as independently coded by the author and a second coder.

[b] Task = task contained *lower familiarity* (school bus) versus *higher familiarity* (toll booth), manipulated between subjects

[c] Counterfactual Mindset = *counterfactual* mindset prime versus *no counterfactual* mindset prime, manipulated between subjects

[d] Example = example audit evidence was present verses absent, manipulated between subjects

TABLE 20
Inferential Statistics Regarding Elaboration of Idea Generation[a]
(Dependent: Elaboration of Proposed Audit Evidence; Factors: Example, Counterfactual Mindset, Task; Covariate: Time)

Panel A: Analysis of Covariance

Source (between subjects)	SS	df	MS	F-Statistic	p-value
Task[b]	0.06	1	0.06	0.01	0.928
Counterfactual Mindset[c]	7.65	1	7.65	1.06	0.305
Example[d]	20.46	1	20.46	2.83	0.095
Task by Counterfactual Mindset	47.20	1	47.20	6.54	**0.006**
Task by Example	3.49	1	3.49	0.48	0.488
Counterfactual Mindset by Example	13.86	1	13.86	1.92	0.168
Task by Counterfactual Mindset by Example	3.20	1	3.20	0.44	0.507
Time	154.68	1	154.68	21.42	<0.001
Error	931.49	129	7.22		

Panel B: Simple Effects Analogous to H1

Simple Main Effects of Counterfactual by Task	SS	df	MS	F-Statistic	p-value
Elaboration of idea generation is greater for counterfactual versus no counterfactual given higher familiarity	46.90	1	46.90	6.50	**0.006**
Elaboration of idea generation is greater for counterfactual versus no counterfactual given lower familiarity	8.50	1	8.50	1.18	0.280

Panel C: Planned Contrast Interaction of Counterfactual and Example (Analogous to H2)

Lower Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is lower the provision of an example will result in the greatest elaboration (+3) versus providing no intervention (-1), a counterfactual mindset prompt in isolation (-1), or both prompts together (-1).	201.05	1	201.05	27.84	**<0.001**

Higher Familiarity	SS	df	MS	F-Statistic	p-value
When familiarity is higher the provision of an example (-1) will result in greater elaboration versus providing no intervention (-3) and prompting a counterfactual mindset will result in the greatest originality regardless of example presence (+2) or absence (+2).	168.52	1	168.52	23.34	**<0.001**

Notes:
Reported p-values are two-tailed unless testing a one-tailed prediction, as signified in bold face.
[a] I measure the elaboration, or the degree of specification of the audit evidence ideas generated, as the sum of the number of points awarded as independently coded by the author and a second coder.
[b] Task = task contained *lower familiarity* (school bus) versus *higher familiarity* (toll booth), manipulated between subjects
[c] Counterfactual Mindset = *counterfactual* mindset prime versus *no counterfactual* mindset prime, manipulated between subjects
[d] Example = example audit evidence was *present* verses *absent*, manipulated between subjects

APPENDIX A: LIST OF INTERVIEW QUESTIONS FROM STUDY 1

Primary Questions

1. Can you describe for me a time when you were auditing an area or a transaction and it was very *easy* for you to determine whether you had collected sufficient information to conclude?

2. Can you describe for me a time when you were auditing an area or a transaction and it was very *difficult* for you to determine whether you had collected sufficient information to conclude?

3. In addition to anything you noted in the paper questions, were there any other key differences between the two situations that made it easier or harder to know when sufficient evidence had been collected?

4. Can you think of a situation in which two individual auditors possessing the same amount and type of evidence might make different judgments about whether this evidence was sufficient?

5. Can you describe for me a time when client management initially presented you with support for an accounting transaction and you sought out additional evidence?

6. Suppose you are told that an auditor did not collect sufficient audit evidence. What are the most likely reasons for this?

7. How would you describe the process of collecting and reviewing evidence for new/unusual/non-routine transactions? For example, would you typically request one piece of evidence at a time?

Supplemental Questions

1. Do you use different criteria to assess the sufficiency of audit evidence when acting as a reviewer as opposed to when you are the preparer?

    a. If so, what are the major differences? Why do you choose this approach?

    b. What queues might you look for that indicate enough evidence has been collected? Do you do this actively (conscious objective of the review)?

    c. What cues might you look for that indicate enough evidence has not been collected? Do you do this actively (conscious objective of the review)?

2. Are there any "rules of thumb" that you use in general to determine when you have collected enough evidence?

    a. Are there times when these general rules are not appropriate? When? How so?

3. Can you think of any situations when collecting additional audit evidence might lead to a lower quality conclusion?

4. Describe each stopping rule and ask if the participant can recall a time that he or she used that type of rule in the field.

    a. Magnitude threshold - A person's belief about the sufficiency of evidence must reach a predetermined threshold before he or she will stop gathering information.

    b. Mental list – The decision maker creates a criteria set or mental list, when all items satisfied he or she stops collecting additional information.

    c. Difference threshold - The decision maker assesses the incremental value of the most recent piece of information acquired and stops when no longer learning anything new.

    d. Representational stability - The decision maker creates a mental model of the task/situation as evidence is collected, when new information no longer causes changes to the model then information collection ceases.

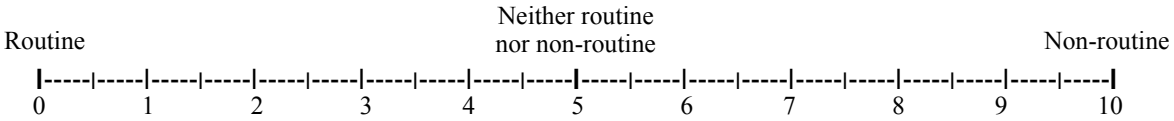
Note 1: The order of questions above is not indicative of the order in which questions were asked during the interview sessions. The interviewer randomized Primary Question 1 and Primary Question 2 by participant as to which scenario (*easy* vs. *difficult*) the participant described first. Participants provided demographic information between describing scenarios.

Note 2: Interviews performed were semi-structured in nature such that the questions above were meant to guide the conversation rather than dictate the discussion points. As such, participants may not have answered every question and the interviewer interjected additional follow-up questions as necessary depending on the topics brought up by the participant within their initial response.

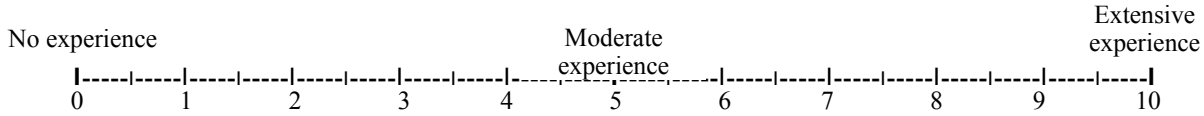
APPENDIX B: SURVEY MATERIALS FROM STUDY 1

Please take a moment to read the following questions and respond by placing an "x" along the scales provided in order to indicate your response. Recall the situation you just described in which it was very easy *(difficult)* for you to determine whether sufficient evidence had been collected in order to conclude. Please characterize this situation based on the various features of an audit task (or audit engagement) provided below.

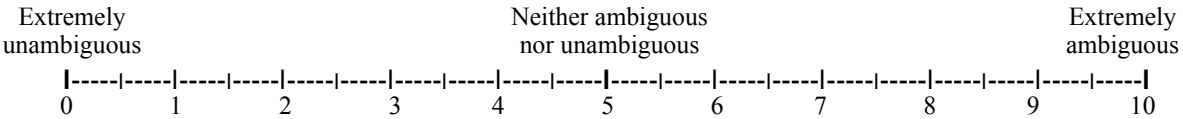
1. How routine was this transaction for the client being audited?



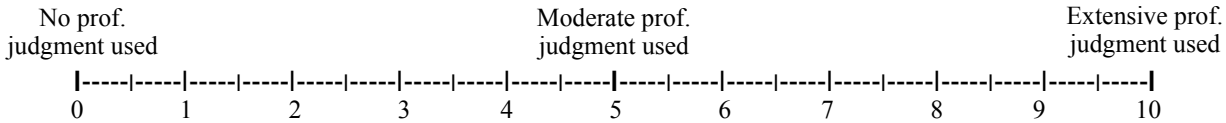
2. How much personal experience did you have with auditing transactions like this?



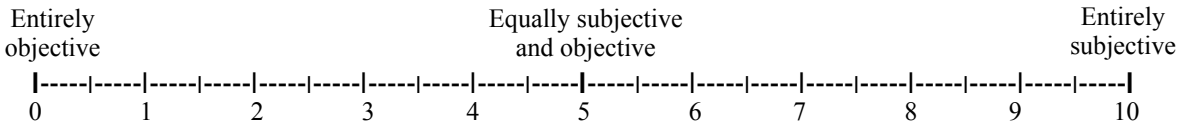
3. To what degree was the applicable accounting guidance unambiguous in that it clearly indicated one correct accounting treatment as opposed to allowing multiple possibly correct methods of accounting?



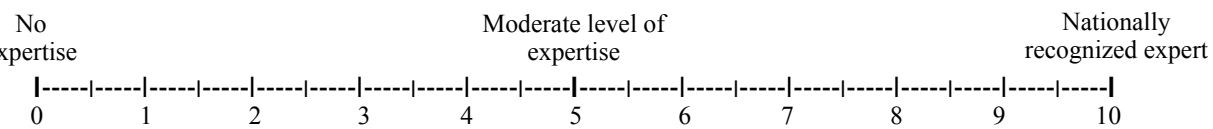
4. How much professional judgment was required to execute the audit program procedures surrounding this transaction?



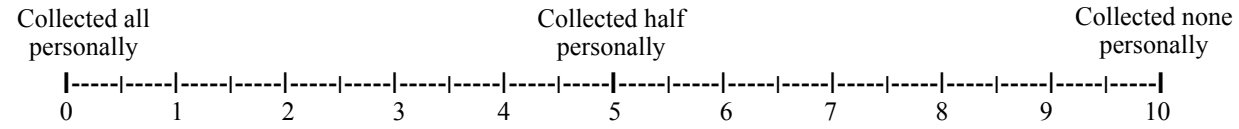
5. On average, how subjective or objective was the evidence that you were reviewing?



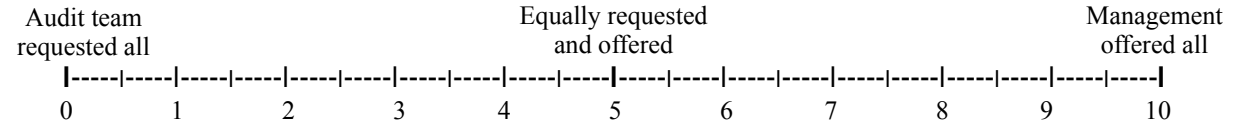
6. For any evidence prepared by other members of your firm, on average what was their level of expertise for this type of transaction?



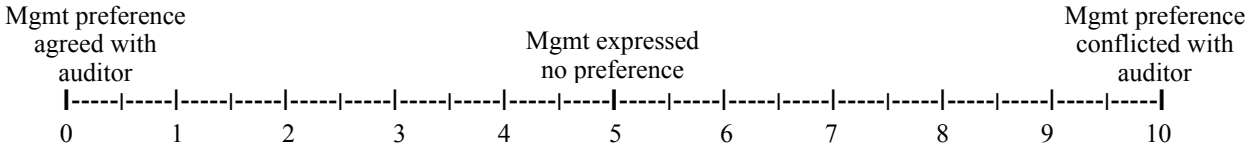
7. To what extent did you personally request and obtain supporting evidence for the transaction/issue?



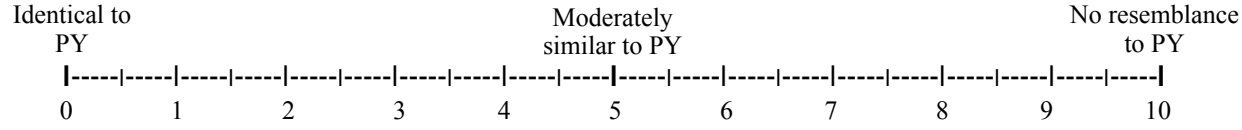
8. To what extent was the evidence reviewed as support for the transaction offered by management as opposed to being received in response to audit team requests for evidence?



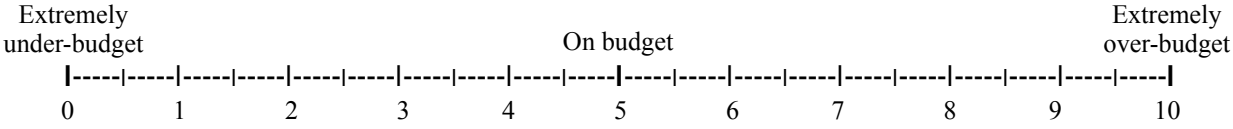
9. To what extent did management express a preferred accounting treatment that either agreed or disagreed with the initial position of your firm?



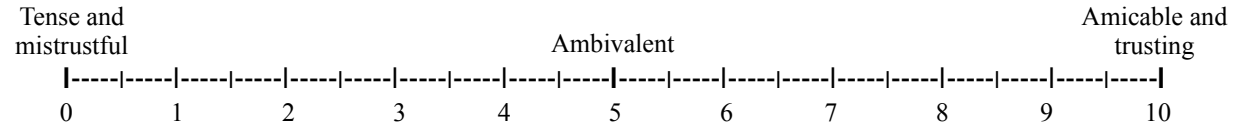
10. How similar was the final working paper to the prior year's working paper?



11. What was the status of the engagement's budget at the time when this issue was reviewed?



12. How would you characterize the relationship of the audit team with the main client contacts?



APPENDIX C: EXPERIMENTAL MATERIALS FROM STUDY 2

Letter of Information and Informed Consent

The purpose of this research study is to better understand how auditors make professional judgments. Your professional experience makes you an ideal candidate to participate in this study. Your participation in this study will take no longer than 40 minutes of your time.

Your participation in this study will involve a brief thought exercise followed by reading a case regarding a hypothetical audit client. After reading the case, you will be asked to respond to questions regarding audit procedures that you and your hypothetical team are performing and the judgments you would make in the described scenario.

Your decision to participate is completely voluntary and you may withdraw from this study at any time without any consequences. While we hope you will be able to answer all questions, you are free to decline to answer any or all questions. You are assumed to have given consent to participate in this study by responding to the questions and returning the completed case to the researchers administering the study. If you decide not to participate, then you may keep the case for your records. Please keep this letter of information and informed consent for your records.

We will not request information about your identity or the identity of your firm, and thus your participation is anonymous. All responses will be entered into a database, saved electronically on secure computers that are only accessible by the researchers, and kept in locked storage. The paper and electronic data will be kept for a period of, at least, five years. The results of this study will be reported in aggregate form, but we might decide to anonymously quote some responses as an illustration of the professional judgments auditors make. These quotes will not be associated in any way with individual participants.

Only the listed researchers will have access to the data during the study. Other researchers may request to view the data later, but they will not know your identity and will not be permitted to publish individual responses.

If you would like further information about the study, please contact Dr. Mark Peecher. If you have any questions about your rights as a participant in this study or any concerns or complaints, please contact the University of Illinois Institutional Review Board at 217-333-2670 (collect calls will be accepted if you identify yourself as a research participant) or via email at irb@illinois.edu. If you have any questions about your or YOUR FIRM's participation in this study, please contact FIRM CONTACT NAME HERE IF YOU WANT or via email at email goes here.

Thank you in advance for your valuable time.

Mark Peecher PhD, Professor
Department of Accountancy
University of Illinois
peecher@illinois.edu
(217)-333-4542

Elizabeth Altiero, Doctoral Student
Department of Accountancy
University of Illinois

General Case Instructions

---

1. There are three parts to this case. Please read the instructions for each part carefully, and then provide your answers to several questions that follow.

2. Please consider your answers carefully before submitting them, as you will not be able to go back and make changes once they have been submitted for research purposes.

3. ***The researchers acknowledge that the information provided in the cases is considerably less than what you normally would have available during an audit.*** Nevertheless, we are interested in your best professional judgment given the available information. Please note that there are no single right answers to the audit case questions.

4. Thank you for your participation.

PART 1

You will first complete a brief thought exercise before proceeding to the audit task. Please read the instructions carefully and respond in the space provided.

<u>Part 1</u>: Thought Exercise

Jane is at a rock concert of one of her favorite bands. Seating is on a first come, first served basis. At the concert the announcer reveals that a trip to Hawaii will be given to a lucky fan and that the winner will be determined by the seat number currently occupied. *[CF Add: Jane's view of the stage is partially obstructed and she sees a much better seat in the near vicinity so she changes seats.]* The announcer returns just prior to the start of the concert to announce the winner of the trip to Hawaii. To Jane's surprise the announcer calls the number of the seat she *[No CF: is sitting in/CF: just moved to]* and she wins the trip to Hawaii.

In the space provided below please write down some examples of thoughts that might run through Jane's head after the concert:

PART 2

Please assume that you are the supervising senior on the client described in the following case. This task will start by providing you with information regarding a hypothetical audit client. Please read this information carefully and respond to the subsequent questions in the spaces provided.

Part 2: Audit Task

[CASE 1] The client is a well-known technology company that often asks job applicants very challenging interview questions. One of these questions asks the applicants, "How many golf balls would fit inside a school bus?"

The legal department of this company is concerned that disgruntled applicants could sue the company if they feel their answers are incorrectly dismissed. Therefore, the lawyers of the firm required that the company purchase a school bus and fill it with golf balls when they started using this question in interviews. Your firm has been engaged to attest as to the number of golf balls currently held inside the school bus.

Your task is to create an audit plan for this engagement by identifying audit evidence you could obtain and audit procedures that you could perform in order to get assurance over the number of golf balls in the school bus.

[Example Condition Add: For example, obtain the dimensions of the school bus and use that information in an analytical procedure to estimate the number of golf balls that the bus can hold.]

[CASE 2] The client is a well-known construction company that is diversifying its operations. In the past year this company purchased a toll bridge that was previously owned and operated by a smaller private company. Toll bridge operations are not an area in which this company has experience, so these operations are completely new to the current year audit.

Your firm has been engaged to attest to the post acquisition financial statements of this company. More specifically, you have been assigned to work on the toll bridge business segment exclusively.

Your task is to create an audit plan for this engagement by identifying audit evidence you could obtain and audit procedures that you could perform in order to get assurance over the revenues associated with the toll bridge's operations.

[Example Condition Add: For example, obtain data on average daily traffic flow across the bridge and use that information in an analytical procedure to estimate the expected annual revenues.]

Please proceed to the following page to record your responses.

1. **What <u>evidence items</u> could you collect and what <u>audit procedures</u> could you perform to complete this engagement?**

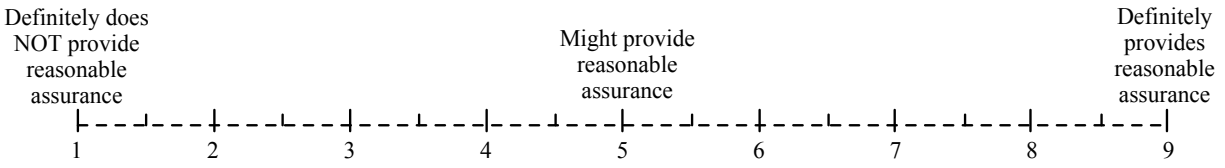
   Please take your time to think of as many ideas as possible. Later you will be using the items listed here to create an audit program so it will be important to have sufficient items to choose from.

2. **Using your responses to Question 1 above, please refine your thoughts into a cohesive audit program that you would recommend using for this engagement.** Write this audit program at a level of detail and specificity such that you would expect a staff on your engagement to be able to execute it. Where necessary, please explain any assumptions that you make.

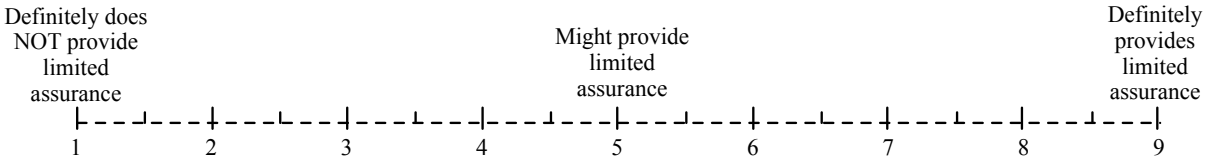
   If you wish to include additional procedures or evidence items that were not included in your response to Question 1 feel free to do so. ***However, please do NOT go back and modify your responses to Question 1.***

3. Please answer the following questions about the program you described on the previous page by placing an "X" on the scale provided.

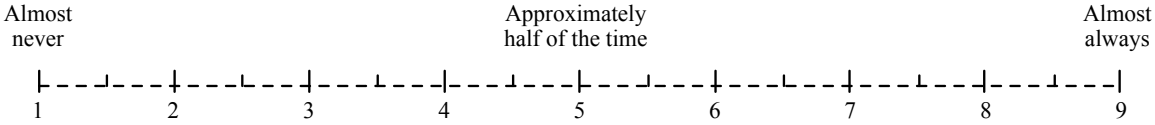
a. To what extent do you believe that testing in accordance with your audit program would provide *reasonable* assurance?



b. To what extend do you believe that testing in accordance with your audit program would provide *limited* assurance?



c. What is the likelihood that your audit program would identify a material misstatement if one exists?

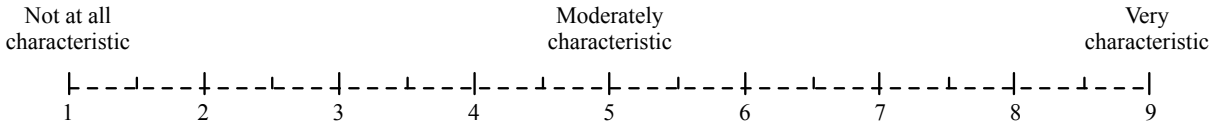


4. Approximately how many total chargeable hours of your time would be required for you to perform the audit program steps specified on the previous page?

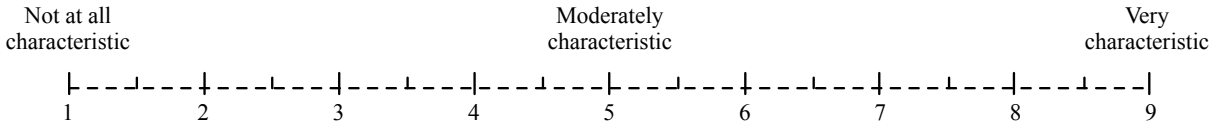
Number of Hours: _____

5. Place an X on the scale provided for each of the following items. To what extent are the following words characteristic of your current mental state:

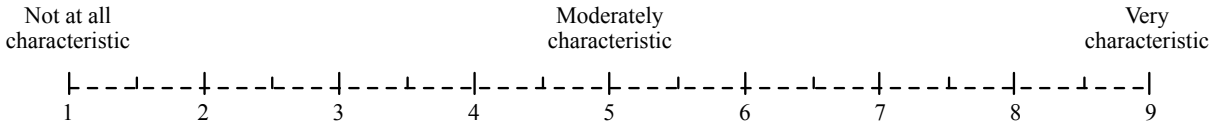
a. Creative



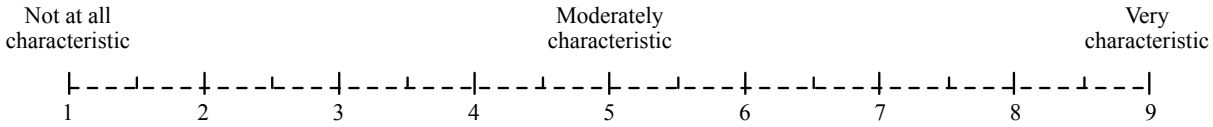
b. Analytic



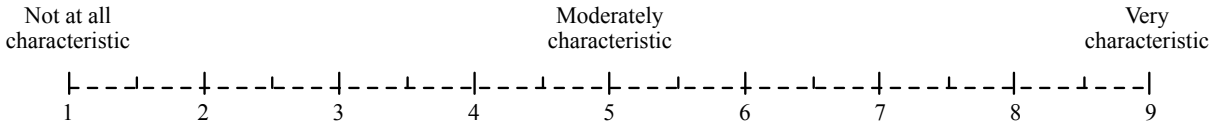
c. Open



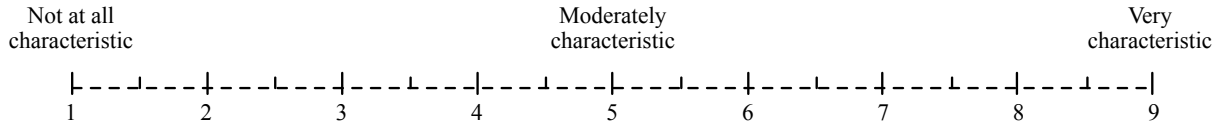
d. Happy



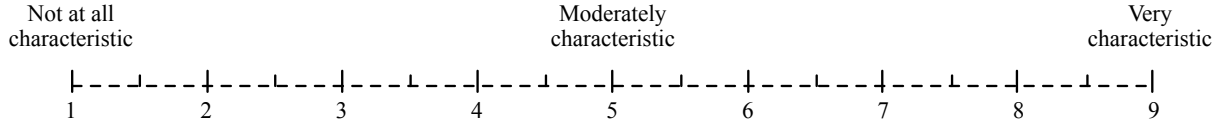
e. Critical



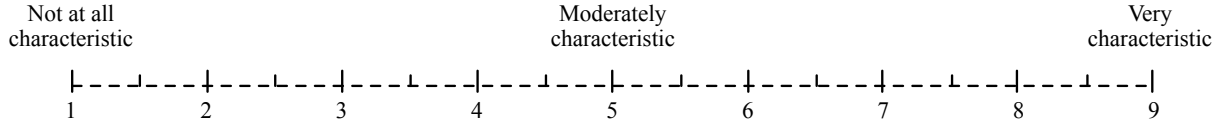
f. Focused



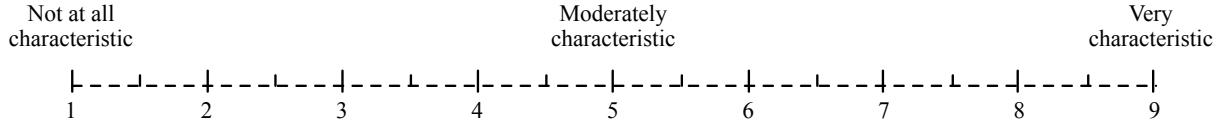
g. Thorough



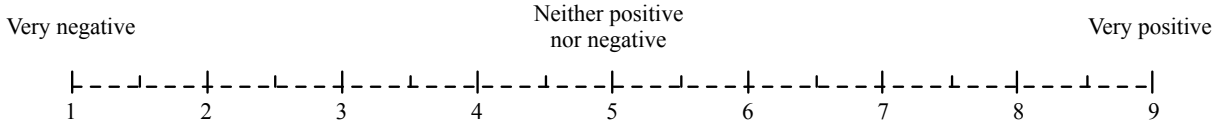
h. Smart



i. Inclusive



j. Mood



You have completed Parts 1 and 2.
Please place this packet into Envelope 1.

Only one more part to go!
Continue the experiment by opening Envelope 2.

PART 3

In this section you will review a brief audit case for a hypothetical audit client and modify an existing audit program for that client's investment account.

Following the audit case you will be asked some general questions about your professional experiences. Please answer all questions to the best of your ability.

Part 3(A): Brief Audit Case

ABC, Inc. is a mid-sized manufacturing client. You are reviewing a potential audit program for the investments account of this client. The company has an effective control environment and its investment portfolio primarily consists of bond and mutual fund holdings.

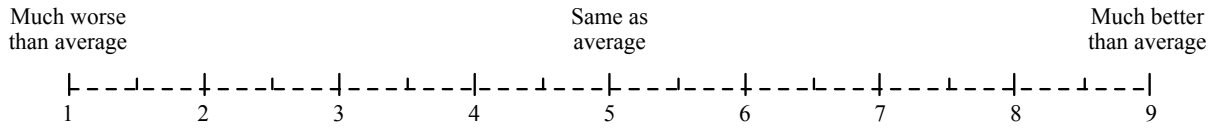
Please review each audit procedure below and indicate whether or not you would *include* the procedure in the audit program. Space has also been provided for you to indicate any *additional* procedures not included on this list that you would add to the audit program.

Investments Account – Audit Program

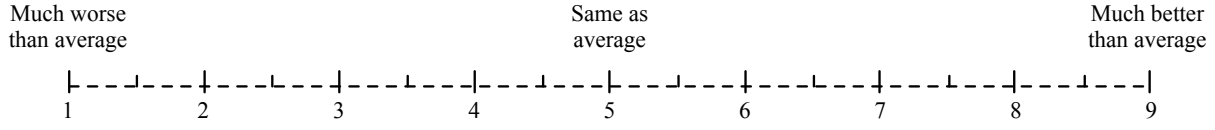
Audit Procedure	*Include Procedure in Current Year Audit Program? (Circle Yes or No)*
1. Perform an analytical review to compare current year investment account balances with prior year's balances.	Yes / No
2. Examine broker's advices for a sample of securities purchased during the year.	Yes / No
3. Search for purchases of securities by examining transactions for a few days after year-end.	Yes / No
4. Review brokers' invoices for cost basis of securities purchased.	Yes / No
5. Determine basis for valuing investments by tracing values to published quotations for marketable securities.	Yes / No
6. Determine whether there has been any permanent impairment in the value of the cost basis of an individual security.	Yes / No
7. Send portfolio information to audit firm valuation specialists for independent verification.	Yes / No
8. Obtain listing of investments by category (held-to-maturity, trading, and available-for-sale); foot listing and agree totals to securities register and general ledger.	Yes / No
9.	Yes
10.	Yes
11.	Yes
12.	Yes
13.	Yes

Please answer the following questions related to the task you completed on the previous page.

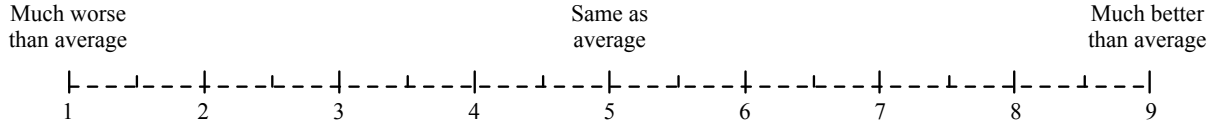
1. *Prior to* any changes you made, how would you rate the **effectiveness** of the eight-step audit program provided?



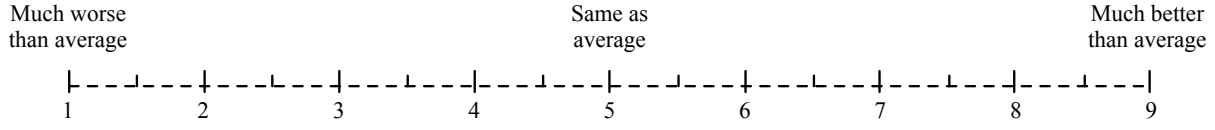
2. *Prior to* any changes you made, how would you rate the **completeness** of the eight-step audit program provided?



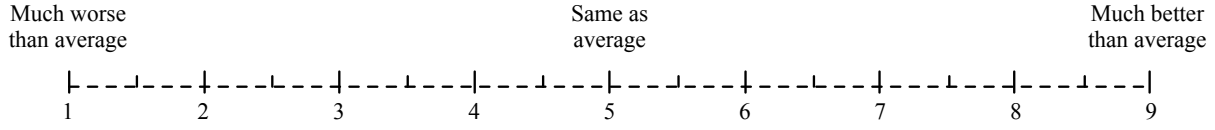
3. *Prior to* any changes you made, how would you rate the **efficiency** of the eight-step audit program provided?



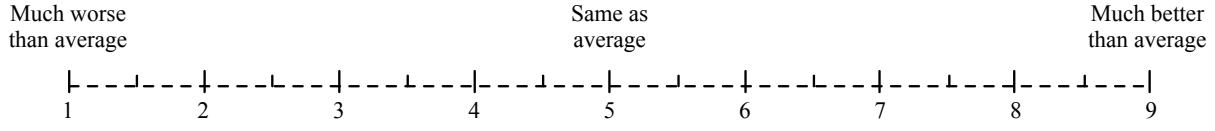
4. *After* any changes you made, how would you rate the **effectiveness** of the modified audit program?



5. *After* any changes you made, how would you rate the **completeness** of the modified audit program?



6. *After* any changes you made, how would you rate the **efficiency** of the modified audit program?



Part 3 (B): General Questions

Please answer the following questions by placing an "X" next to the best answer.

1. In the story you read about Jane attending a rock concert, Jane switched seats just before winning the trip to Hawaii.

_____ True

_____ False

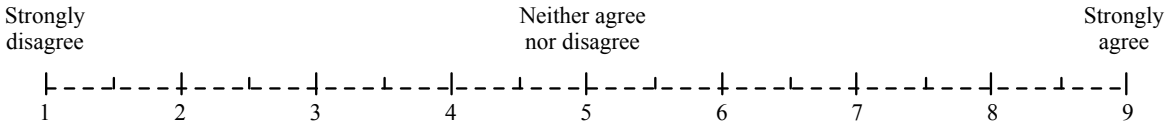
2. The instructions preceding the audit task included the following example audit procedure and related evidence: *[SB: Obtain the dimensions of the school bus and use that information in an analytical procedure to estimate the number of golf balls that the bus can hold./TB: Obtain data on average daily traffic flow across the bridge and use that information in an analytical procedure to estimate the expected annual revenues.]*

_____ True

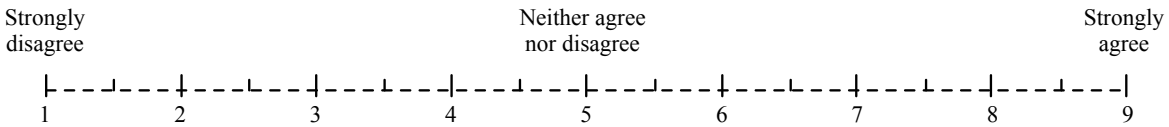
_____ False

Please answer the following questions by placing an "X" on the scale provided

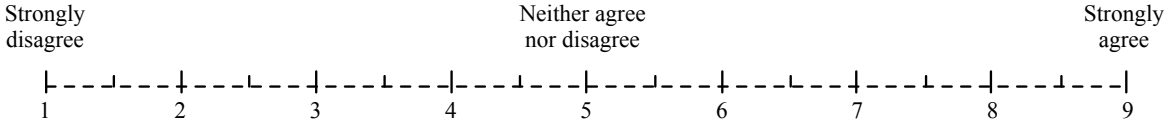
3. I feel confident about my performance on the audit task contained in these materials.



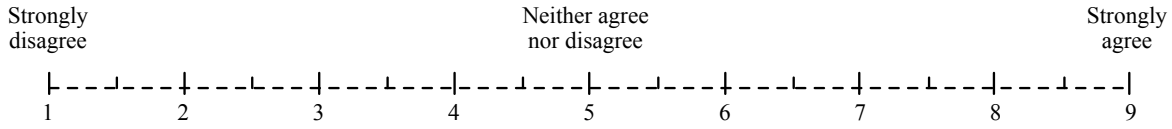
4. In general, I am confident in my ability to make valuable modifications to the audit program when necessary.



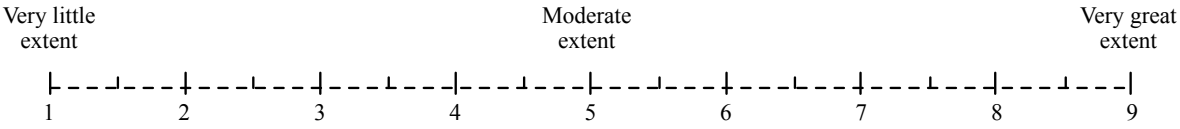
5. I found this task interesting and engaging.



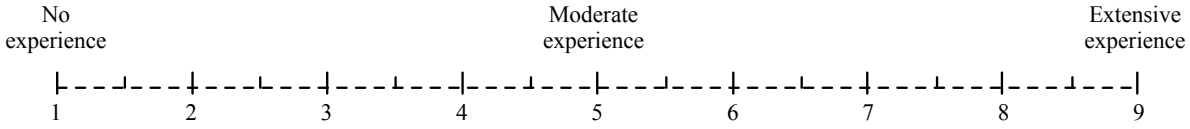
6. In general, I find designing new audit procedures interesting and engaging.



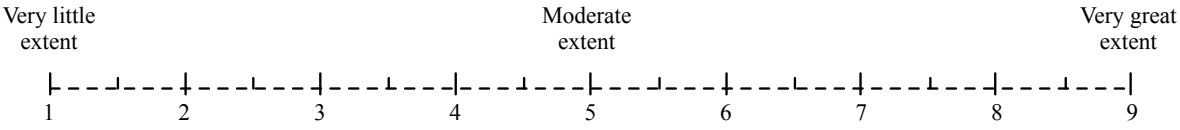
7. To what extent to you agree with the following statement: I often encounter situations in which I must modify an existing audit program.



8. How much experience do you have in **designing** or **modifying** audit procedures?



9. For audit procedures you have **designed** or **modified** in the past, to what extent do you believe that they added value to the audit?



10. How many months of audit experience do you have?

Number of months: _____

11. What is your industry of specialization? (If you do not have a designated industry of specialization respond please respond "N/A.")

_____

12. What is the most creative audit procedure you have ever designed or witnessed during your time as an auditor? Please describe briefly below:

Almost done! For the final four questions, please think about the experiences that you have had with current and former clients.

1. Which members of management have had responsibility for making asset impairment decisions, for accounting purposes? (Check all that apply.)

_____CEO

_____CFO

_____COO

_____Other (please specify)_____

2. Which members of management have had responsibility for making future capital investment decisions? (Check all that apply.)

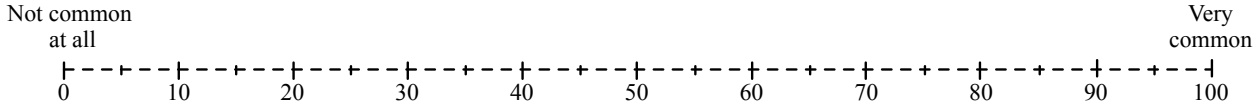
_____CEO

_____CFO

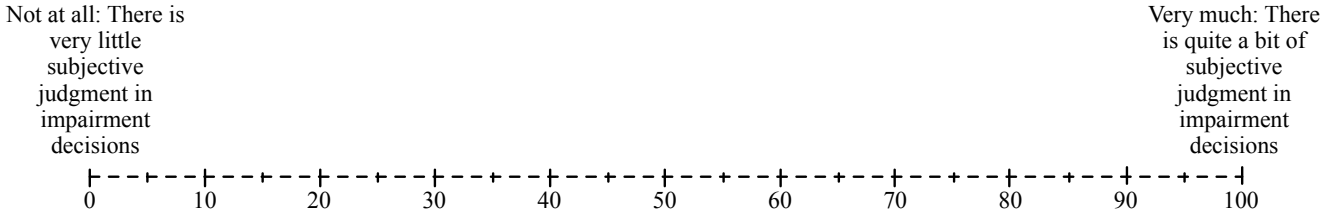
_____COO

_____Other (please specify)_____

3. In your experience, how common is it for the same member of management to have responsibility for BOTH an asset impairment decision as well as future capital investment decisions related to that asset?



4. In your experience, how subjective are management impairment decisions?



THANK YOU FOR PARTICIPATING!
Please return all materials to the last envelope and return the entire packet to the experiment administrator.