The Effect of Staff Auditor Reputation on Behavior Related to Audit Quality

Emily S. Blum  
Culverhouse School of Accountancy  
The University of Alabama  
Esblum1@crimson.ua.edu

Richard C. Hatfield*  
Culverhouse School of Accountancy  
The University of Alabama  
rhatfiel@cba.ua.edu  
(205) 348-2901

Richard W. Houston  
Culverhouse School of Accountancy  
The University of Alabama  
rhouston@culverhouse.ua.edu

February 2020

* Corresponding author

Acknowledgements: We acknowledge Lori Bhaskar, Heather Carrasco, Ben Commerford, Mary Kate Dodgson, Sean Hillison, Patrick Hurley, Bill Messier, Mike Ricci, Arnie Wright and workshop participants at the Workshop on Audit Quality (EAA), University of Kentucky, Indiana University, NEBARS, Northeastern University, and the Norwegian School of Economics for their helpful comments. We also thank the following individuals for their help in administrating the experiments: Seanna Asper, Chris Blum, Mariel Knight, James Lawson, Chris McCoy, Julie Mercado, Ranier Robinson, and Daniel Street.
The Effect of Staff Auditor Reputation on Behavior Related to Audit Quality

Abstract

Auditors often face situations in which acting on their professional obligations comes with potential personal costs. Drawing upon the Theory of Reputation in Organizations as well as Social Presentation Theory, we predict and find that the perceived costs associated with these actions are lower for auditors with a positive reputation, and that these perceptions influence their actions. In our first experiment, we find that auditors with positive reputations are more likely to proactively report expected budget overages and less likely to engage in dysfunctional behavior than those with negative reputations, who perceive a more constrained choice set. In a second experiment, we manipulate reputation with a supervisor in a live-action simulation and demonstrate a causal link between reputation and auditors’ skeptical action. Taken together, our experiments provide evidence that an auditor’s perceived reputation influences their choice to engage in behavior related to audit quality, independent of their actual competence.

**Keywords:** reputation, audit quality, information sharing, skeptical action
I. INTRODUCTION

Staff auditors often encounter situations in which they face competing goals and motivations, such as when taking skeptical action bothers clients (e.g., Bennett and Hatfield 2013), speaking up about an audit issue irritates a supervisor (e.g., Nelson and Proell 2018), or exceeding the budget hurts their evaluation (Agoglia, Hatfield, and Lambert 2015). In this study, we examine how staff auditors’ personal reputation within their firm can influence the actions they take when they find themselves in a situation where performing an audit quality enhancing action (QEA) could come with personal costs. Such costs include social conflict (e.g., upset manager or client) as well as a more direct cost of poor performance evaluation, which can cause specific career limitations (e.g., salary, promotions, client selection). We draw upon the Theory of Organizational Reputation as well as Social Presentation Theory to predict and find that auditors with positive (negative) reputations expect better (worse) evaluations when they engage in QEAs. Auditors with low reputations are more likely to expect good evaluations if they engage in dysfunctional behavior (i.e., underreporting time). In these situations, and as a result of these differing expected costs, auditors with high reputations are more likely to perform QEAs while low reputation auditors are more likely to engage in dysfunctional behavior.

The Theory of Organizational Reputation (Zinko, Ferris, Blass, and Laird 2007) describes reputation as a perceptual identity distinct from, though clearly related to, actual ability or competence, based on the perceptions of others within an organization. One’s reputation can be perceived as different, and even conflicting, by different individuals or groups within the organization (Ferrist et al. 2003). A positive reputation generally provides power for individuals within an organization (Pfeffer 1992). This power is a social power, granted by others within the organization (Zinko 2009) whereby they defer to the individual regarding the topic (Hackman
and Oldham, 1975). This perception of an individual relates to Hollander’s (1958) model in that those with reputations for high competence are given the benefit of the doubt and freedom to act, or *idiosyncrasy credits*. Based on this model, behavior consistent with group norms improves reputation (earns credit) that can later be used (Hollander 1958), for example, to behave skeptically in difficult situations. However, certain audit behavior such as client questioning or going over budget can hurt an auditor’s reputation (i.e., expend credit). Consequently, the power from reputation is temporal because others constantly update, from subsequent direct observation or secondary sources, their perceptions of the individual (i.e., as they earn or expend idiosyncrasy credits), and the power from reputation can in fact be viewed as fragile (Ferris et al. 2003).

The audit environment is one where staff auditors are, and should be, mindful of their reputation. We propose that characteristics such as multiple supervisors, frequent evaluations, clear organizational norms, and comparison to others (e.g., utilization rates, chargeable time) create an environment in which auditors’ reputations impact their personal incentive/cost structure and thereby influence their behavior. Auditors are cognizant of these dynamics. For example, auditors recognize that managing reputation can increase the likelihood of being promoted and assigned to more desirable clients (e.g., Guénin-Paracini, Malsch, and Paillé 2014). However, the social dynamics of the audit environment may cause a misalignment between personal reputation goals and a firm’s audit quality goals.

Auditing research asserts that auditors are frequently placed in difficult situations where social pressures (e.g., Bennett and Hatfield 2013) or individual professional costs can discourage QEAs. For example, Brazel, Jackson, Schaefer, and Stewart (2016) demonstrate how the outcome effect causes supervisors to penalize staff who take appropriate action to investigate
potential fraud when no fraud is subsequently discovered. Related studies suggest that avoiding certain QEAs (e.g., Commerford, Hatfield, Houston, and Mullis 2017; Nelson and Proell 2018) or engaging in dysfunctional behavior (Agoglia et al., 2015) can result in more positive outcomes for staff auditors. Applying theory to the audit context, auditors with positive reputations incur in lower personal costs when they conduct QEAs, such that QEAs are more likely. Similarly, auditors with negative reputations are less likely to be given the benefit of the doubt, and thus more likely to rely on dysfunctional behavior to improve their outcomes (i.e., performance evaluations). That is, auditors with positive vs. negative reputations have different cost structures when choosing whether to engage in QEAs or other actions, which predictably alters behavior.

We designed two experiments with complementary methods to examine how reputation affects staff auditors’ personal cost incentive structure (i.e., relative evaluations) and their resultant likelihood of engaging in QEAs or dysfunctional audit behaviors.1 In Experiment 1, practicing auditors consider the actions of a young staff auditor, who has developed either a positive or negative reputation based on his actions on three prior engagements. We also consider a third condition, which separates reputation from competence, in which the auditor has a negative reputation with a particular manager, but has exhibited high competence on other engagements. Because the first manager is unaware of the staff auditor’s subsequent highly competent work, participants infer that he has a negative reputation with that manager. We use this condition to isolate the effect of reputation by portraying a staff auditor with a negative reputation who actually is highly competent. The QEA in this experiment is proactively telling the manager that a task will go over budget, which may irritate the manager (see Nelson and

---

1 All experiments were approved by the institutional review board at the affiliated university.
Proell 2018), but would allow the manager to incorporate the overage into the audit plan and preserve audit quality. The dysfunctional behavior of interest is underreporting time.

Results suggest that auditors with a positive reputation expect to receive favorable evaluations despite going over budget when they engage in a QEA whereas, in the other conditions, auditors do not expect to receive a favorable evaluation if they engage in a QEA and exceed budget. Without a positive reputation, auditors view dysfunctional behavior (i.e., underreporting time) as the only way to receive a favorable evaluation. We also conduct a mediation analysis to demonstrate that expected evaluations (i.e., personal cost) alter the likelihood of engaging in both QEAs and dysfunctional behavior.

In Experiment 2, student participants engage in simulated audit tasks in which we focus on QEAs to demonstrate a direct link from our reputation manipulation to auditor behavior. We manipulate participants’ reputation via an in-person supervisor’s evaluation—the supervisor (an actor) sets either low or high expectations and stresses either positive or negative aspects of participants’ work. After receiving this evaluation, participants perform a simulated inventory count and walk-through, with opportunities to engage in QEAs by requesting explanations and asking for supporting documentation for observed issues. Such follow-up on issues imposes personal costs because it results in the task taking longer (going over budget) and delays client work in the warehouse (the client became increasingly irritable as participants increased the delay through information/evidence requests). Participants in the positive reputation condition

---

2 We discussed this issue with a group of highly experienced current and former auditors (i.e., manager through partner level from firms ranging in size from Big 4 to large local, and with experience ranging from six to 40 years), who stated that speaking up about overages early allows them to understand and respond when the audit plan needs revision. It also allows them to make resource adjustments—if teams do not speak up about overages in a timely manner, then the team may have to complete the audit with constrained resources, which threatens audit quality.
request more explanations and documentation (i.e., engage in QEAs by taking skeptical actions) than do the participants in the negative reputation condition.

Our findings contribute to the recent audit literature on QEAs such as employee voice and skeptical action (e.g., Nelson, Proell, and Randel 2016; Gissel and Johnstone 2017; Kadous, Proell, Rich, and Zhou 2019) and dysfunctional behavior (e.g., Agoglia et al., 2015). Our results demonstrate that perceived personal reputation is an important influential link between audit behavior and associated personal costs. Where previous studies have demonstrated that auditors perceive personal costs associated with increased audit effort (e.g., Nelson et al. 2016, Bennett and Hatfield 2013), we demonstrate how reputation causes different auditors to perceive these costs differently, which ultimately influences their intended actions.

We also extend work on reviewer bias. Where previous research has found that audit supervisors display reputation-related bias in performance evaluations (Tan and Jamal 2001), we find that subordinates not only recognize and anticipate supervisor bias, they also change their behavior accordingly. This tacit knowledge could be leveraged by firms to reduce the perceived costs of QEAs. Given that these costs are somewhat entrenched (Agoglia et al. 2015; Brazel et al. 2016), this is potentially a more fruitful route than attempting to force behavior through firm policy or regulations (Bazerman, Loewenstein and Moore 2002). Specifically, auditors can perhaps be nudged toward optimal behavior through consideration of their reputation. This is consistent with the management style approach discussed in Nelson et al. (2016). It is promising that consideration of these costs of behavior can empower auditors to engage in costly positive behavior even in a situation where there is unlikely to be a personal benefit.

Finally, this study advances the reputation literature by demonstrating a feedback loop theorized by prior research (Zinko et al. 2007; Zinko et al. 2012). Specifically, we demonstrate
how reputational outcomes (i.e., the benefit of the doubt) influence the reputation building process (i.e., self-presentation). In our setting, reputation affects choice of behavior, which is explained by differences in anticipated supervisor evaluations. Thus, we provide empirical evidence that having a positive reputation makes it easier to maintain that reputation through QEAs. Our results are particularly interesting as, in contrast to much of the prior literature, the action of interest is undertaken not for personal benefit, but for the benefit of audit quality.

II. HYPOTHESIS DEVELOPMENT

Personal Reputation within an Organization

An individual’s reputation within an organization is not solely a function of their competence, but also a subjective social construct that is perceptual by nature (e.g., Fine 1996; Gamson, Croteau, Hoynes, and Sasson 1992; Rao 1994). A common definition of reputation within an organization, from Zinko et al. (2007) is as follows: “a perceptual identity formed from the collective perceptions of others, which is reflective of the complex combination of salient personal characteristics and accomplishments, demonstrated behavior, and intended images presented over some period of time as observed directly and/or reported from secondary sources, which reduces ambiguity about expected future behavior” (165). It is also important to note that within an organization an individual can possess different, and even conflicting, work-related reputations perceived by different individuals or groups (Ferris et al. 2003).

Reputations within an organization influence several potential outcomes, including professional outcomes such as promotion, salary increases, and strong evaluations, as well as more personal outcomes (e.g., a better home life and happiness; Zinko et al. 2012). Reputations are built by behavior that is consistent with the expectations and norms of the organization and are impaired by behavior that goes against these norms (Ferris et al. 2003). Below, we will
discuss how an auditor’s management of their reputation incents different behaviors depending on their current perception of their reputation.

**Reputation Dynamics in Audit Firms**

Public accounting firms possess key contextual characteristics in unique combinations that make workplace reputations particularly salient to their staff. Accounting firms conduct frequent evaluations, possess easily comparable benchmarks (e.g., utilization rates), and hold clear expectations/norms (e.g., time budgets, workpaper completeness, positive client relationships). All of these features are important aspects of reputation development (e.g., Ferris 2003). Additionally, the audit environment causes staff auditors to work with multiple supervisors, and supervisors to work with multiple staff auditors. These clear benchmarks, coupled with frequent comparisons that come with changing teams, make reputation an easy and important piece of information for decision-makers.

Evidence suggests that reputation-related information is used frequently in audit settings. Those with direct experience with staff readily share their perceptions with others within the firm, including those who determine engagement staffing (Guénin-Paracini et al. 2014; Westermann, Bedard, and Earley 2015). Further, Tan and Jamal (2001) find that previous impressions of a subordinate (i.e., reputation) influence audit managers’ assessments of subordinate work quality. Thus, a strong professional reputation is a key contributing factor to staff auditors’ success.

Audit staff understand that their reputation is a key factor in their success. A senior associate interviewed by Guénin-Paracini et al. (2014) states: “You get very quickly labelled in an auditing firm. […] [For example if] your superior has it in for you, his friends ask him how you’re doing, he pulls you to pieces, and little by little you get a reputation as a numskull (sic).
To avoid getting caught up in something like that, you’ve just got to do everything you can to avoid getting poor assessments.” (277) While the quote suggests that poor assessments can result in negative reputations, auditors also perceive that, while quality work is important, a positive reputation is required as a way to “play the game” (i.e., being placed on top clients; Life of an Auditor 2010).

Given the importance of reputation in the audit environment, as well as the salience of reputational issues, it is likely that auditors will be cognizant of their reputation (Hochwater et al. 2007), and that this knowledge will influence their actions. However, awareness of reputation as a key factor for success will likely affect personal incentives, such that auditors may be dissuaded from actions consistent with audit quality goals. Consistent with this incentive, Smith, Hutton, and Jordan (1996) find that auditors perceive that underreporting time – which is considered a dysfunctional audit behavior – will yield better evaluations, higher perceptions of their competence, better assignments, and career advancement.

**Reputation and Power to Act**

A positive reputation generally provides power for individuals within an organization (Pfeffer 1992). Zinko (2009) suggests that, because reputation is a social construct, the power it produces is not legitimate power, which is granted by the organization through position (i.e., title). Most closely linked to our setting is expert power, a power given by others in the organization, whereby they defer to the power-holder regarding the topic of their expertise (Hackman and Oldham, 1975).³ This perception of an individual relates to Hollander’s (1958) 

---

³ Related to status power is reward/coercive power defined by Keltner et al. (2003) as “an individual’s capacity to modify others’ states by providing or withholding resources or administering punishments.” While this type of direct power influences behavior similar to our predictions, through approach versus inhibition tendencies, the source of power is quite different.
model, in that those with reputations for being highly competent are given the benefit of the
doubt and freedom to act, which Hollander refers to as idiosyncrasy credits.

Importantly, idiosyncrasy credits are accrued as individual behavior conforms to a social
group’s expectations, and expended as individual behavior deviates from these norms (Hollander
1958). Thus, the power arising from reputation builds as an individual meets and exceeds
expectations, but diminishes as the individual deviates from those expectations. Consequently,
the power from reputation is temporal due to constant updating from subsequent direct
observation or secondary sources of the individual’s behavior (i.e., as they earn or expend
idiosyncrasy credits), and can in fact be viewed as fragile (Ferris et al. 2003; Herbig, Milewicz
and Golden 1994).

Reputation and Audit Quality Enhancing Actions

A common situation in which auditors face a choice of whether or not to deviate from
social group expectations is the consideration of audit quality enhancing action. While
organizations and professional standards emphasize the importance of audit quality, the literature
is clear that there are salient social group pressures to avoid QEAs, as they often involve
“rocking the boat” by requiring more time, deviation from planned procedures, or risking client
relationships. For example, Nelson and Proell (2018) find that audit supervisors may be irritated
when, during the audit, staff raise audit issues that could increase audit effort, although they
often reward staff for voicing issues in subsequent performance reviews. Brazel et al. (2016) find
that supervisors impose penalties when staff undertake additional investigation based on
observing a potential issue when that effort did not identify a misstatement. Evidence also
suggests that supervisors, on favored clients, evaluate their staff less positively if they charge all
of their time and exceed budget, rather than underreporting time (Agoglia et al. 2015).
Consequently, it appears that perceived overzealousness or exceeding budgets result in idiosyncrasy costs (i.e., annoyed supervisor and/or poor evaluations). It is important to note that these costs can be fairly entrenched in a firm’s culture and difficult to eliminate despite partner preference and firm guidance (Agoglia et al. 2015). In summary, QEAs often come with associated personal, social, and professional costs, including the salient costs caused by deviating from social norms and expectations. We consider how a strong reputation may change the perception of these costs, leading to more QEAs and less dysfunctional behavior.

Employees are motivated to manage the impressions of others (Bozeman and Kacmar 1997) using the levers that are most controllable (Ferris and Judge 1991) to demonstrate conformity with, or excellence relative to, the organization’s expectations and norms (Ferris et al. 2003). Social Presentation Theory indicates that evaluations within an organization serve as integral components of the construction of one’s public image (Baumeister 1982). Thus, in difficult audit situations (i.e., those performing QEAs may bear personal costs), an auditor with a weak reputation is in the position of needing to earn credit by conforming with expected norms. Alternatively, when reputation is strong, the auditor is in a better position to go against norms (i.e., spend credit). Assuming that auditors prefer to fulfill their professional obligations by conducting a quality audit, as long as the costs are not too high, we expect auditors with strong positive personal reputations to choose QEAs even when they involve deviating from norms.

It is important to note that supervisors who evaluate staff are unlikely to notice a staff’s failure to engage in a particular QEA (e.g., Bennett and Hatfield 2013); consequently, when staff do not perform QEAs, their reputation is unlikely to be harmed; in fact, their reputation could even be enhanced. For example, failure to more fully investigate an issue could help staff avoid negative feedback and also help them meet budget (i.e., earn credit) (e.g., Brazel et al. 2016).
Similarly, engaging in dysfunctional behavior can result in earned credit (Agoglia et al. 2015). For auditors, reputation is a particularly relevant construct in their environment, and staff recognize it as a key factor for success (Guénin-Paracini et al. 2014; Life of an Auditor 2010).

In summary, meeting budget, keeping clients happy, and not bothering supervisors are observable outcomes that conform to expectations within the audit firm. If a situation arises where these norms must be violated to engage in QEA, then auditors with more positive reputations are expected to perceive the cost of such violations as lower than auditors with less positive reputations. We assert that auditors with weaker reputations will perceive higher costs associated with QEAs, and they will therefore emphasize managing their reputation by meeting these norms, seeing greater value in engaging in dysfunctional behavior. Given these differing cost structures for these two behaviors, auditors with strong reputations will be more likely to engage in QEAs (i.e., spend credit) and less likely to engage in dysfunctional behavior (i.e., earn credit) than those with weaker reputations.

H1a: Staff auditors with stronger reputations will anticipate better evaluations for engaging in QEAs than auditors with weaker reputations.

H1b: Staff auditors with weaker reputations will anticipate better evaluations for engaging in dysfunctional behavior than for engaging in QEAs.

H2a: Staff auditors with stronger reputations are more likely to engage in QEAs than auditors with weaker reputations.

H2b: Staff auditors with weaker reputations are more likely to engage in dysfunctional behavior than auditors with stronger reputations.

III. METHOD

We conduct two experiments to test the hypotheses. In the first, we employ a third person methodology to examine how auditors believe a staff auditor will behave in a specific situation. This experiment allows us to address both sets of hypothesis while considering multiple possible
behaviors. The second experiment focuses on H2a and involves a controlled setting where we manipulate reputation among our participants to more directly examine the causal link between reputation and auditors’ engagement in QEAs.

**Experiment 1**

**Participants**

Participants include alumni of a large state university in the Southeast. We sent a targeted recruitment e-mail to students who had graduated from the Masters of Accountancy program in the last 1-3 years, and also posted recruitment messages on the department’s social media sites.\(^4\) We offered an incentive to participate: we told potential participants that, for each completed survey, we would donate $25 to a scholarship fund for current students. Over a ten-day period, we received 98 completed responses: 64 from e-mail and 34 from social media.\(^5\) Sixty-eight percent of respondents were currently employed as an auditor, and the rest had until very recently been employed as a full time staff auditor.

**Experimental task**

The experimental task, conducted on a computer-based platform, provided descriptions of Jack, a new staff auditor at a large international audit firm. To date, Jack had worked on three engagements with three different supervisors.\(^6\) We manipulate Jack’s reputation by describing aspects of Jack’s performance on each engagement. Our 1 x 3 design considers three conditions—positive, negative, and negative\(^\text{false}\) reputation. After the manipulation, we measure

\(^4\) Beyond the benefit of access to these auditors, the recruitment of participants having a common background also reduces unimportant cross-sectional differences. However, we also acknowledge the potential limitation related to generalization.

\(^5\) The e-mails were sent to approximately 250 alumni, yielding a 27.2% response rate by e-mail. We do not have data on how many social media users viewed our posts, but the department’s largest social media account, its Facebook account, had 1,518 followers at the time we wrote this footnote (many of whom are current students).

\(^6\) We use a third person vignette because it allows us to manipulate reputation through a series of past performances, and study unethical behavior (e.g., underreporting hours, skipping steps) while limiting social desirability bias.
participants’ perceptions of Jack’s reputation with his supervisor on his upcoming fourth engagement, and then describe the details of the fourth engagement. It is on this fourth engagement that participants evaluate the likelihood that Jack will take one of a number of possible actions (including actions beneficial or detrimental to audit quality).

**Independent Variable**

We manipulate Jack’s reputation by varying descriptions of his performance on his first three engagements, each of which had different supervisors, as well as through Jack’s discussion with his “coach” (see Appendix A for a full description of the manipulations). We specifically focus on Jack’s reputation with his first supervisor, Sam. For the three engagements in the positive reputation (PR) condition, Jack uses technology to increase efficiency and meet budget, finds and addresses an error with minimal oversight and meets budget, and asks insightful questions on a plant tour, respectively. In the negative reputation (NR) condition, Jack struggles with technology and goes over budget, fails to uncover an error and then requires guidance to address the error and goes over budget, and fails to wear appropriate attire for a plant tour, resulting in his inability to attend the tour and complete his assignment.

Recall that the reputation construct is perceived rather than objectively measured. The PR and NR manipulations demonstrate Jack’s reputation with his supervisor, but likely also send signals about his competence, which would affect participants’ evaluations of Jack’s likely future actions independently of reputation. To separate reputation from competence, we include a third condition, Negative\textit{false} (NFR), in which Sam perceives Jack to have a negative reputation, but where Jack showed improvement and demonstrated strong competence. In the NFR condition, we describe Jack’s first engagement (where Sam is the manager) identically to the NR condition, but we describe the next two engagements identically to the PR condition. In both the NR and
NFR conditions, Jack has a negative reputation with Sam; however, in the NFR condition, Jack demonstrated competence on subsequent engagements. That is, in the NFR condition, Sam only knows of Jack’s poor performance on his first engagement and not his subsequent improvement. In the PR and NR conditions, however, Sam knows of Jack’s performance on all three engagements, allowing us to isolate the effect of perceived reputation from the effect of competence. Analyses demonstrate that participants viewed Jack as competent in the NFR condition, which provides tension for our hypotheses in that we can demonstrate whether it is actual or perceived competence (i.e., reputation) that influences perceptions of the likelihood of QEAs or dysfunctional behavior.

After participants view the summary of the first three engagements, they view a summary of Jack’s check-in with his “coach,” who is responsible for overseeing Jack’s professional development. This meeting summarizes Jack’s performance on the first three engagements from a secondary source to reinforce the reputation manipulation, and also to reinforce Jack’s reputation with Sam (client 1 supervisor), who will be his supervisor on the fourth engagement.

**Dependent Variable**

The QEA of interest is participants’ judgments about whether Jack will proactively notify the supervisor about a potential budget overage on the fourth engagement (H1a, H2a). Doing so requires a proactive discussion with the supervisor suggesting a revised budget, and is considered a “voice” behavior (e.g. Nelson et al. 2016; Nelson and Proell 2018), in that it involves speaking up about a problem to a supervisor with a goal of operational improvement. Voice behaviors carry risk as they involve challenging the status quo, which may damage relationships (Van Dyne and LePine 1998; Detert and Bruno 2017). In our setting, the voice behavior challenges the budget set by the supervisor by suggesting that the time budget underestimates the actual time an
average associate would take to complete the task. Proactively warning the supervisor is a QEA as it could enhance audit quality by helping the supervisor manage staff time and anticipate overages. Timely notification helps the audit team avoid resource constraints when additional work or staffing is needed to complete the audit, which poses threats to audit quality if not done. However, it comes with potential personal costs in that the supervisor could react poorly to the notification (Nelson and Proell 2018). The dysfunctional behavior of interest is underreporting time (H1b, H2b). Underreporting time is considered an unethical practice and has downstream effects that threaten audit quality (Ponemon 1992; Donnelly, Quirin, and O’Bryan 2003; Sweeney and Pierce 2006). It is the likely way that audit staff can conduct all appropriate work and meet budget in situations like the one described in this experiment (Agoglia et al. 2015).

We introduce a situation where, on his fourth engagement, Jack realizes that he likely will go over budget. Participants view four options Jack can take in response to this predicament, and provide anticipated supervisor evaluations for each action individually. They consider two options in which Jack exceeds the budgeted hours: one in which Jack proactively informs Sam of the expected overage, and one in which Jack completes the work over budget without first informing Sam of the overage. Participants also consider two additional potential actions. Participants provide Sam’s anticipated evaluation if Jack completes the task and does not charge all hours worked, which is considered a dysfunctional behavior. Participants also consider the evaluation received if Jack rushes through testing (i.e., engages in over-documentation, also known as ghost or phantom ticking), a dysfunctional behavior that directly threatens quality

---

7 We discussed the implications of proactively speaking up about budget overages with a group of experienced current and former auditors. They noted a number of operational costs to delayed notification, including an impact to audit quality that can arise from timing and resource constraints. One former auditor with over 30 years of experience summarized the discussion as follows: “In theory, it [finding out about an overage later] shouldn’t matter, you do what you have to do to what it takes [to complete a quality audit]. But that’s not the real world.”
(Sweeney and Pierce 2006). Participants respond on an 11-point scale to the statement “please assess how you believe Sam will rate Jack’s performance in his performance review, on a scale from 1 (unacceptable) to 11 (exceptional),” with the midpoint labelled as “acceptable.” We chose this midpoint to the evaluation scale, based on the finding in Brazel et al. (2016) that auditors view meeting expectations as a below average result, such that auditors would need to be above this midpoint to receive a positive review. Participants also assess the likelihood that Jack would choose each of the four possible actions by allocating a percentage likelihood (out of a total of 100%) to each action.8

**Results**

**Manipulation Check**

To measure Jack’s reputation with Sam, participants rate their agreement with five statements adapted from Hochwarter, Ferris, Zinko, Arnell, and James’ (2007) reputation scale, which focuses on technical reputation (Cronbach’s alpha = 0.983). Participants in the PR condition rated Jack’s reputation with Sam as significantly higher than in the NR condition (means of 9.79 vs. 2.04, t=29.64, p < 0.0001) and in the NFR condition (mean of 9.79 vs. 3.16, t = 21.41, p < 0.0001). While participants in the NFR condition rated Jack’s reputation with Sam as higher than those in the NR condition (mean of 3.16 vs. 2.04, t = 3.35, p = 0.0014), the mean reputation rating for Jack in the NFR condition was still below the scale midpoint of 6 (t = -10.44, p = < 0.0001, one-tailed), suggesting that we effectively manipulated a poor reputation with Sam in both the NR and NFR conditions.

---

8 We partially randomize the order in which the options are presented: the two within-the-rules options (no action, proactive notification) are always presented before the two outside-the-rules options (not charge all hours worked, documenting work not performed), but are randomized within those blocks. We made this choice to mirror a decision process in which unethical options are considered only after ethical options. As all choices are provided to participants before any evaluations are made, participants evaluate a complete choice set when anticipating supervisor responses. No order effects were evident in our hypothesis tests.
Hypothesis Tests

H1a predicts that auditors anticipate better outcomes for positive-reputation auditors engaging in a QEA than those with negative reputations. We first conduct a two-way mixed ANOVA with reputation (between-subjects) and option (within-subjects) as independent variables (untabulated), and find significant main effects for both reputation (between subjects, $F = 12.27, p < 0.001$) and option (within-subjects, $F = 69.99, p < 0.001$) on anticipated performance rating, with no significant interaction.

When Jack engages in a QEA (i.e., proactively informs Sam of budget overage), participants in the PR condition anticipate higher performance ratings from Sam than those in the NR condition ($t = 3.82, p = 0.0002$, one-tailed) and in the NFR condition ($t = 4.92, p < 0.0001$, one-tailed), strongly supporting for H1a. The difference between the NR and the NFR conditions is not significant ($t = 1.07, p = 0.30$). We also compare anticipated performance ratings to the scale midpoint (i.e., “acceptable”). Because a positive reputation is built on positive deviations from norms and expectations (Zinko et al. 2007), to obtain or maintain a positive reputation, auditors must exceed what is “acceptable.” We find that participants in the PR condition anticipate success with proactive notification (i.e., their performance ratings exceed “acceptable;” $t = 7.18, p < 0.0001$, one-tailed). In contrast, those in the NFR and NR conditions do not anticipate positive evaluations ($t = 1.24, p = 0.89$, one-tailed, and $t = 0.22, p = 0.41$, one-tailed, respectively).

To test H1b, we compare the evaluations for the NFR condition depending on whether Jack engaged in a QEA or underreporting. Consistent with H1b, NFR auditors received a higher evaluation when underreporting time than when engaging in a QEA ($t_{29} = 2.97, p = 0.003$, one-tailed). It is interesting to note that, in all conditions, participants anticipate positive outcomes
when Jack underreports his time to meet the budget. Reassuringly, participants do not anticipate positive outcomes if Jack skips steps. In summary, participants anticipate that, while all auditors (even those without a positive reputation) can obtain positive personal outcomes (i.e., evaluation above acceptable) by underreporting time such that they meet the budget, they anticipate that only those auditors with a positive reputation will obtain a positive outcome by engaging in a QEA.

[Insert Table 1 here]

H2a predicts that the likelihood that staff auditors will engage in a QEA is contingent upon reputation. Participants believe that the PR auditor is more likely to proactively inform his supervisor of the budget overage than the NR auditor (t = 3.32, p < 0.001, one-tailed) and marginally more likely than the NFR auditor (t = 1.36, p = 0.090, one-tailed). Generally, PR auditors are more likely to report all their hours, regardless of proactive notification, than NFR (t = 1.83, p = 0.036, one-tailed). Consistent with H2b, PR auditors are expected to be less likely to engage in dysfunctional behaviors (i.e., underreport hours) to meet budget than NR (t = 2.15, p = 0.018, one-tailed) or NFR auditors (t = 2.24, p = 0.014 one-tailed).

[Insert Table 2 here]

We also consider whether the differences in action likelihoods are attributable to the different expectations of evaluations for these actions. To do so, we conduct a mediation analysis in which we use the 5-item reputation scale because it allows us to determine how participants’ perceptions of Jack’s reputation influence their perceptions of Jack’s actions. We find that anticipated performance ratings from engaging in a QEA fully mediate the positive relationship between Jack’s reputation with Sam and the likelihood of engaging in the QEA. We also find that anticipated performance ratings mediate the negative relationship between Jack’s reputation
with Sam and the likelihood of underreporting time. Bootstrapped confidence intervals for both indirect effects are significant at a 99% confidence level, consistent with our prediction that the differential personal cost (benefit) of a QEA (dysfunctional behavior), due to auditor perceived reputation, influences auditor behavior during an audit.

[Insert Figure 1 here]

**Additional Analyses**

Recall that the NFR condition separates reputation from competence, as Jack performed poorly on the first engagement, but performed well on subsequent engagements. However, the supervisor on the original audit (Sam) only knows of Jack’s work quality on that engagement. Consequently, while we describe an auditor with generally high competence, he has a poor reputation with Sam. In hypothesis tests, participants in the NFR condition respond similarly to those in the NR condition, indicating that it is the negative reputation, not perceptions of competence, that causes differences between the NFR and PR conditions.

We asked additional questions about Jack to evaluate further whether participants perceive Jack as competent in the NFR condition, despite his negative reputation with Sam. First, we ask how long an “average associate” would take to complete the Client 4 task, on an 11-point scale anchored at -5 (significantly less than Jack), 0 (same as Jack), and +5 (significantly more than Jack). Participants in the NR condition believed that an average associate would take less time than did Jack (t = 5.82, p < 0.0001, one-tailed), whereas participants in the NFR and PR conditions did not (p-values > 0.5; non-tabulated), suggesting that participants in the NFR condition perceive Jack as competent.

We also asked participants general questions about Jack to understand Jack’s reputation with the *participant*. Participants in the NFR condition were more likely than those in the NR
condition to seek advice from Jack both about technical issues ($t = 8.99, p < 0.0001$), and about problems with supervisors ($t = 5.0, p < 0.0001$). Although participants perceived Jack to be more competent in the NFR condition than in the NR condition, they expect him to be equally unsuccessful in obtaining positive evaluations for positive behavior, and thus equally likely to engage in dysfunctional behavior. Consequently, the NFR condition provides evidence that Jack’s reputation is driving the findings about Jack’s actions, not Jack’s competence.

**Experiment 2**

Experiment 1 shows that participants perceive that auditors are more likely to engage in positive (dysfunctional) behavior when they have a positive (negative) reputation. Experiment 2 focuses on QEAs and uses a $1 \times 2$ between-participants design that extends and replicates these findings by measuring *participant’s actions* (as opposed to perceptions of someone else’s actions) in an audit simulation. That is, we create a situation in which participants are placed in an audit predicament (i.e., whether or not to bother the client) and are concerned with their supervisor’s evaluation. Experiment 2 also extends Experiment 1 by considering another QEA, professional skepticism.

**Participants**

As a proxy for staff-level auditors with minimal experience, we use as participants 52 Masters of Accountancy students who average 2.9 months of audit experience. Participants (50% female) are from a large state university in the Southeast.

**Experimental Task**

Participants completed two audit tasks, designed for entry level staff capabilities, in individual rooms with laptop computers that had links to case materials: accounts receivable testing and inventory observation. Before completing the tasks, participants met an actor playing
the role of audit supervisor, which allowed the participant to know that an actual person would be providing feedback on their work, and to meet their supervisor before they received feedback. Participants were told that their work would be reviewed and that they would receive both written and verbal feedback, similar to feedback from supervisors in practice.

Participants first completed an accounts receivable (AR) confirmation testing case, in which they tested confirmations and performed alternative procedures. The case included a single seeded exception: a confirmation was received, but the customer responded that the balance was incorrect. If participants chose to inquire further, they would obtain evidence supporting the client’s balance, not the customer’s response. After approximately 17 minutes had passed, the supervisor came to their room and performed an “in-process review.” This in-process review served as the implementation of the reputation manipulation.

We then asked participants to perform an inventory observation task. This task involved a computer-simulated inventory count where participants observed and counted inventory items and received audio explanations in response to their inquiries. Before beginning this task, participants answered questions to measure their perceived reputation with the supervisor. After a brief tutorial, participants completed the inventory observation: they move through five “sections” of the warehouse, with each section including a sheet-to-floor count, a floor-to-sheet count, and an observation. Counts contained minor seeded variances and each observation involved a description of an issue observed while in the warehouse. For example, one observation identified some damaged items that had been set aside and excluded from the count. Participants were given the option of whether or not to inquire further. If they chose to inquire further, they received an explanation from the warehouse manager, and they were then given the
option to request support or not. When participants finished the case, they completed questions about the case and provided demographic information.

**Independent Variable – Reputation with Actor**

We manipulate participant reputation by varying the interaction with the actor playing the role of “Sam,” the supervisor. The AR case varied the extent to which it allowed participants to complete the task in a timely manner. Participants in the positive reputation condition could complete more selections than those in the negative reputation condition. While all participants viewed the same audit support and completed the same tests in the same order, participants in the negative reputation condition (NR) tested multiple invoices to complete a single alternative procedure, while participants in the positive reputation condition (PR) used those same invoices to complete multiple alternative procedures. We pilot tested the AR case using a group of undergraduate students. At the 17-minute mark, pilot participants had completed between 3 and 11 testing selections. We used this information to create supervisor expectations.

We manipulated reputation by setting expectations for both efficiency and effectiveness, and then having participants either exceed or fall short of such expectations. The full feedback script is provided in Appendix B. Efficiency expectations were set as three selections for the positive reputation group, and 10 selections for the negative reputation group, ensuring that most PR participants would have completed more than the benchmark, but NR participants would have completed fewer. For the effectiveness portion of reputation, the reviewer provided feedback on the participant’s documentation of the confirmation that was received with an exception. Software was installed on participant computers that allowed the researchers, together with the actors, to view participant documentation before the actor performed the review. We chose between three types of specific feedback (commenting on the explanation of the exception,
the documentation of the investigation, or the conclusion) based on actual participant
documentation. In the PR condition, the reviewer praised a particular element of the
participant’s documentation, whereas in the NR condition the reviewer suggested improvements
on a particular element of the participant’s documentation.10

**Dependent Variable**

The dependent variable is the level of skeptical action in the Inventory Observation task,
which we operationalize as the number of times the participant requests either more information
or supporting documentation about an observation. The case consists of five observations (see
Appendix C) and, for each observation, a participant can request more information and then
choose whether to request supporting documentation; consequently, the dependent variable can
range from 0 to 10. All observations are purposefully ambiguous as to whether an exception
exists—no evidence provided should lead the participant to believe that inventory quantities are
potentially materially misstated; however, some evidence may be relevant to other audit areas.
Client responses are provided in audio clips where the supervisor provides the requested
explanation. It is important to note that these requests alter the warehouse manager’s behavior—
he becomes a bit more irritated and comments that these requests increase the likelihood that the
inventory count will not be done in time to prevent work disruption.

---

9 It is important to note that there were no qualitative differences between conditions with regard to *actual* efficiency
or effectiveness. That is, there were no actual competence differences between conditions. The manipulation
involved providing different expectations and stressing positive or negative aspects of documentation.
10 Given the unique design (i.e., altering how participants believe somebody perceives their abilities), we provided a
short debrief to those participants in the negative reputation condition immediately after they completed the
experiment. We told them that the actor was purposefully concentrating on negative aspects of their work to
simulate a critical supervisor. After all participants finished the experiment, we had a larger debrief that included a
broad discussion of their performance (e.g., managing the client) and discussion of the resulting behavior. Written
feedback was also provided at that time.
Results

We designed the materials intending for every participant to correctly identify the exception in the AR testing, as the actor reviews the participant’s work on this issue to manipulate reputation. However, 17 participants (32.7%) did not identify the exception. Given that we designed the exception to be obvious, we presume that missing this exception suggests a lack of adequate effort or audit knowledge. More importantly, we are unable to adequately manipulate reputation if participants did not provide the basis for doing so. We therefore remove these 17 participants—our analyses include only the remaining 35 participants.11

We base our reputation manipulation check measure on a seven-question scale adapted from Hochwarter et al. (2007). Our manipulation alters participants’ beliefs about how their supervisor perceives their audit ability. Participants responded to the questions (e.g., “Sam thinks I am a very competent auditor,” “I have a good reputation with Sam”) by stating their agreement on seven-point scales (Cronbach’s alpha = 0.904). Participants in the high-reputation condition perceived their reputation with Sam to be significantly higher than those in the low-reputation conditions (means 5.66 vs. 3.56; p-value < 0.001), suggesting that the manipulation was effective.

Recall that the outcome variable of interest is the extent of skeptical action, specifically the number of requests for explanations and supporting documentation from the warehouse manager. Participants made more requests in the positive reputation (4.6) compared to the negative reputation condition (3.1; p = 0.02, one-tailed).12 We also analyzed the type of request

---

11 Prior research suggests that if participants are inattentive to the case, results become obscured (Maniaci and Rogge 2014). Consistent with this notion, if these 17 participants are included in the analyses, they significantly alter our findings in a way that justifies omitting their responses (the reputation manipulation had no effect).
12 There were two confederates who played the role of supervisor. Based on a marginally significant interaction when confederate is added to the model, one of the confederates has a marginally stronger effect on the dependent variable than the other. If we look at the simple main effect of reputation for the weaker confederate, the dependent variable is in the same direction, but not significant (n = 16).
and found that positive reputation participants were more likely to inquire further than were negative reputation participants (3.00 vs. 2.33, \( t = 1.79, p = 0.042 \) one-tailed) and were more likely to request additional documentation (2.33 vs. 0.80, \( t = 1.96, p = 0.029 \)). Post-experimental questions suggest that, relative to negative reputation participants, positive reputation participants liked Sam more (5.90 vs. 4.93; \( p = 0.042 \), two-tailed) and expected better evaluations from Sam (5.05 vs. 3.73; \( p = 0.002 \), two-tailed). Consistent with expectations, positive reputation participants were more concerned with getting the right answer for the inventory case than were negative reputation participants (1.45 vs. 2.13; \( p = 0.005 \)).

[Insert Table 3 here]

Experiment 2 provides a direct causal link between reputation and auditor skeptical behavior by controlling for actual competence through random assignment. Taken together, Experiments 1 and 2 support our prediction that QEAs (e.g., reporting an error, speaking up about an audit issue) carry risk that differs by the type of reputation, thereby affecting whether an auditor will act in ways that improve audit quality.

IV. Discussion

In this study, we report the results of two experiments that test whether an auditor’s perceived personal reputation influences their willingness to act in a way that contributes to overall audit quality, but may involve some personal risk. Experiment 1 demonstrates that auditors perceive the personal cost of engaging in a QEA when faced with an anticipated budget overage to be less when the auditor has a positive reputation with his supervisor than when the auditor has a negative reputation with his supervisor, causing positive reputation auditors to be more likely to engage in a QEA. Further, auditors with positive reputations are less likely to engage in dysfunctional behavior than auditors with negative reputations. For negative reputation
auditors, dysfunctional behaviors were the only way to receive positive evaluations, demonstrating a differential cost structure for positive vs. negative reputation auditors.

While our results indicate that we successfully isolated reputation from competence in terms of its influence on action, Experiment 2 confirms these results by using an audit simulation instead of a hypothetical vignette. Thus, our dependent variable is an observed action with real social outcomes in the simulated audit case. Experiment 2 also extends our findings to another type of QEA: skeptical action. We find that students, as proxies for novice auditors, engage in more skeptical actions when they have a more positive reputation with a supervisor than when they have a less positive reputation.

Our findings significantly and meaningfully move the reputation literature forward. Reputation researchers have theorized that the outcomes of reputation can influence the reputation-building process, creating a feedback loop (Zinko et al. 2007; Zinko et al. 2012). Specifically, Zinko et al. (2012) posit that reputation may change the definition of “deviation from norms.” In Study 1, we find that reputation changes the characterization of the budget overage: participants use reputational cues to conclude that the average associate would take less time than Jack only when Jack had an earned negative reputation. This, in turn, eases the path to additional reputation-building (in terms of receiving positive evaluations) for those with a positive reputation. By demonstrating how perceptions of existing reputation can affect future reputation-building actions, we provide empirical evidence that this feedback loop exists.

Our findings also contribute to the audit literature. We build on the work by Tan and Jamal (2001) and place it in a broader social psychological framework. While they find that audit supervisors display reputation-related bias in evaluations, we demonstrate that subordinates not only recognize and anticipate supervisor bias in evaluations—they also change their behavior
because of it. This social awareness is a type of tacit knowledge, which is valued in audit contexts as it allows auditors to navigate socially interactive tasks (Bol, Estep, Moers, and Peecher 2018). However, our finding highlights a potential negative consequence of tacit knowledge, in that it may cause auditors to limit QEAs in certain social contexts.

We also contribute to a growing audit literature on the personal costs associated with QEAs. Audit researchers have previously identified various risk concerns that can affect an auditor’s willingness to act in ways that affect audit quality: social mismatch affects skeptical action (Bennett and Hatfield 2013), leadership type affects willingness to speak up about audit issues (Nelson et al. 2016), and error climate affects error reporting (Gold, Gronewold, and Salterio 2014). Some have proposed that such auditor behaviors are influenced by the incentive structures that these auditors face (e.g., Agoglia et al. 2015), and indeed all of these studies share the assumption that some actions benefitting audit quality carry some potential personal cost. In addition, further studies have demonstrated that these costs exist: supervisors impose penalties for unfruitful skeptical actions (Brazel et al. 2016), they get irritated when subordinates raise issues (Nelson et al. 2018), and they sometimes prefer when staff do not report budget overages (Agoglia et al. 2015). We highlight an important boundary condition – perceived personal reputation – which can influence how audit staff anticipate their supervisors will respond to their decision to increase audit effort, thus affecting their decisions about how to act.

We argue that the features of the audit environment, such as team switching, clear benchmarks, and the use of subordinate auditors as valued but interchangeable resources, increase the salience of reputation dynamics. Future research could study whether auditors who work in smaller offices, where team switching is more common, are more cognizant of reputation dynamics than auditors who work in large offices and spend the bulk of their time on
one large client. It is likely that the development of this tacit knowledge is facilitated by team switching, making the audit environment a unique setting to study reputation.

This study draws from the reputation and power literatures to motivate the hypotheses. However, there is an important distinction between these findings and the assorted theories related to power – we study a subordinate’s power over a superior, whereas much of the power literature focuses either on structural power (superior power over a subordinate), or power dynamics within a group of peers. We demonstrate how reputation leads auditors to anticipate better outcomes when reporting an anticipated budget overage, demonstrating that they anticipate a better ability to exert influence over their supervisor. Again, the audit environment presents a unique context to study subordinate power, as audit associates are relatively interchangeable, but a high-quality subordinate is a valuable asset. Future research in organizational behavior could leverage this environment to develop theories of subordinate power.

Future research can explore further how an individual’s personal reputation influences the character and interpretation of risks in the audit setting. Additionally, research can consider how firm tone at the top and supervisor managerial approach can alter both the construction of staff auditor reputation, through defining expectations, as well as more directly influencing the incentive structure of the auditor. While we focus on a comparison between positive and negative reputation, our theory would also apply to situations where auditors have not yet accumulated sufficient reputational capital – such as when auditors first start work, or transfer to a new team or office. Given the inevitability of reputation dynamics, researchers and practitioners can respond to these findings by considering theory-based strategies supervisors and employers can use to empower auditors with negative or less-developed reputations to take action towards audit quality.
References


Appendix A (Experiment 1)
Reputation Manipulations

We first develop Jack’s reputation through a discussion of his performance on previous clients, in which he deviates either positively or negatively from expectations. In the changed reputation condition, he deviates negatively from expectations on the first engagement, but positively on the second two engagements. This creates a positive overall reputation, but a negative reputation with Sam.

<table>
<thead>
<tr>
<th>Changed Reputation</th>
<th>Negative Reputation</th>
<th>Positive Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same as Negative</td>
<td>On Client 1, Jack worked inefficiently, though effectively, on all assigned tasks. He spent a lot of extra time figuring out how to use the firm's technology, ultimately completing all tasks, but not in a timely manner. His senior, Sam, was disappointed with Jack's inefficiency and technological skills. Overall, Sam noted that Jack performed worse than the average audit associate.</td>
<td>On Client 1, Jack worked efficiently and effectively on all assigned tasks. He used the firm's technology to complete all tasks in a timely manner, exhibiting superior skill. His senior, Sam, was pleased by Jack’s efficiency and technological skills. Overall, Sam thought that Jack performed better than the average audit associate.</td>
</tr>
<tr>
<td>Same as Positive</td>
<td>While working on Client 2, Jack failed to detect an error in the audit area to which he was assigned. Jack failed to detect the error, in part, because of his weak technological skills. After his senior, Blake, found the error during his review of Jack's work, Blake determined the additional testing necessary, brought the error to Jack's attention, and asked Jack to perform the additional testing and document the error. Blake was disappointed that Jack failed to detect the error, needed a lot of oversight to address the error after Blake discovered it, and exceeded the budgeted hours for the work he completed.</td>
<td>While working on Client 2, Jack uncovered an error in the audit area to which he was assigned. Jack detected the error, in part, because of his strong technological skills. After briefly consulting with his senior, Blake, about the error and his plans to address the error, Jack performed additional testing and documented the error. Jack completed his work on the audit area under budget despite the additional work he performed to detect and document the error. Blake was impressed that Jack found the error on his own, addressed it effectively with minimal oversight, and still managed to come in under budget.</td>
</tr>
<tr>
<td>Same as Positive</td>
<td>While working on Client 3, Jack's audit team was scheduled to tour the client's manufacturing plant. Prior to the tour, Jack's supervisor Pat had provided him with instructions on what to expect during the tour, including appropriate attire, and had also given him a list of points to consider and questions to ask while on the plant tour. Jack's main responsibility during the tour was to take notes on discussions between Pat and the plant manager. On the day of the tour, Jack forgot to wear appropriate attire, and had to stay behind in the audit workroom while the rest of the team went on the plant tour. Pat was disappointed that Jack did not wear appropriate attire even though Pat had told him about the required attire ahead of time.</td>
<td>While working on Client 3, Jack's audit team was scheduled to tour the client's manufacturing plant. Prior to the tour, Jack's supervisor Pat had provided him with instructions on what to expect during the tour, including appropriate attire, and had also given him a list of points to consider and questions to ask while on the plant tour. Jack's main responsibility during the tour was to take notes on discussions between Pat and the plant manager. On the day of the tour, Pat was impressed with Jack's appropriate attire, Jack's detailed and accurate notes, as well as Jack's ability and willingness to ask important questions during the tour.</td>
</tr>
</tbody>
</table>
Appendix A (Experiment 1)
Reputation Manipulations - Continued

We next reinforce Jack’s overall reputation via a third-party source: Jack’s coach. We also reinforce Jack’s reputation with Sam. In the changed reputation condition, participants are informed that Sam has not heard about Jack’s performance on engagements other than Client 1.

<table>
<thead>
<tr>
<th>Changed Reputation</th>
<th>Negative Reputation</th>
<th>Positive Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In their meeting, Jack’s coach said the following about Jack’s experience so far:</td>
<td>In their meeting, Jack’s coach said the following about Jack’s experience so far:</td>
<td>In their meeting, Jack’s coach said the following about Jack’s experience so far:</td>
</tr>
<tr>
<td>“I’ve heard from Sam, Blake, and Pat about your performance. Sam told me he was</td>
<td>“I’ve heard from Sam, Blake, and Pat about your performance. They all were not</td>
<td>“I’ve heard from Sam, Blake, and Pat about your performance. They all were very</td>
</tr>
<tr>
<td>not impressed, but Blake and Pat told me that they were very impressed. While Sam</td>
<td>impressed. They all told me that your performance fell below their expectations for</td>
<td>impressed. They all told me that your performance exceeded their expectations for</td>
</tr>
<tr>
<td>told me that your performance fell below his expectations for an audit associate at</td>
<td>an audit associate at your level. You are off to a weak start, but can recover by</td>
<td>an audit associate at your experience level. You are off to a strong start, but will</td>
</tr>
<tr>
<td>your experience level, Blake and Pat told me that your performance exceeded their</td>
<td>doing good work going forward. It looks like you will be working with Sam again on your</td>
<td>need to keep up the good work. It looks like you will be working with Sam again on</td>
</tr>
<tr>
<td>expectations for an audit associate at your experience level.</td>
<td>next engagement. I hope the engagement goes well.”</td>
<td>your next engagement. I hope the engagement goes well.”</td>
</tr>
<tr>
<td>You have recovered strongly from your weak start, but will need to keep up the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good work. It looks like you will be working with Sam again on your next engagement.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I hope the engagement goes well.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>For his fourth engagement, Jack again will be working with Sam. Jack knows that Sam</td>
<td>For his fourth engagement, Jack again will be working with Sam. Jack knows that Sam</td>
</tr>
<tr>
<td></td>
<td>has not heard about his engagement performance on Client 2 with Blake and on Client</td>
<td>has heard about his engagement performance on Client 2 with Blake and on Client 3</td>
</tr>
<tr>
<td></td>
<td>3 with Pat. Sam met with Jack prior to the engagement to discuss Sam’s expectations</td>
<td>with Pat. Sam met with Jack prior to the engagement to discuss Sam’s expectations of</td>
</tr>
<tr>
<td></td>
<td>of Jack. During that meeting, Sam said the following:</td>
<td>Jack. During that meeting, Sam said the following:</td>
</tr>
<tr>
<td>Sam: “Your overall performance with me on Client 1 fell short of my expectations.</td>
<td>Sam: “Your overall performance so far has fallen short of expectations, including your</td>
<td>“Your overall performance so far has exceeded expectations, including your work with</td>
</tr>
<tr>
<td>This engagement will be difficult, but it will be an opportunity for you to show</td>
<td>work with me on Client 1. This engagement will be difficult, but it will be an</td>
<td>me on Client 1. This engagement will be difficult, but it will be an opportunity</td>
</tr>
<tr>
<td>what you can do.”</td>
<td>opportunity for you to show what you can do.”</td>
<td>for you to show what you can do.”</td>
</tr>
</tbody>
</table>
Appendix B (Experiment 2)
Reputation Manipulation Script

After the participant has completed approximately 17 minutes of the AR task, the actor re-enters the room to conduct an in-process review. Actors followed the script below, adapting their commentary on the participant’s documentation to address that participant’s work. The manipulation creates a positive or negative deviation from expectations in two categories: efficiency and effectiveness.

<table>
<thead>
<tr>
<th>Review – PR</th>
<th>Review – NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi {{Name}}, we’re about up for our time on this first task. Why don’t you finish up the item you’re working on, and when you get to the review screen I will do a quick review of your AR testwork so far. <strong>A number of students have only completed three sections at this point.</strong> How many have <strong>you</strong> finished?</td>
<td></td>
</tr>
<tr>
<td><strong>(Student responds.)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Wow, that’s much more than I expected!</strong></td>
<td></td>
</tr>
<tr>
<td>In general, when you first start working in practice, most students have the same technical knowledge from school. Efficiency is one way to really stand out.</td>
<td></td>
</tr>
<tr>
<td><strong>You did well on that!</strong></td>
<td></td>
</tr>
<tr>
<td><strong>OK, let me take a look at your documentation. You’ve done great here so far.</strong> Can you walk me through this selection here?</td>
<td></td>
</tr>
<tr>
<td><strong>(Student shows confirmation and explains. Say “good” at appropriate pauses.)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Documenting errors can be challenging. You’ve done a good job of explaining the exception/documenting the investigation/concluding on exception.</strong></td>
<td></td>
</tr>
<tr>
<td>Having the technical skills to write clear conclusions is another way to stand out at work.</td>
<td></td>
</tr>
<tr>
<td><strong>Again, you’ve done well on that!</strong></td>
<td></td>
</tr>
<tr>
<td><strong>OK, let me take a look at your documentation. You’ve done OK here so far.</strong> Can you walk me through this selection here?</td>
<td></td>
</tr>
<tr>
<td><strong>(Student shows confirmation and explains. Say “ok” at appropriate pauses.)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Documenting errors should be straightforward. You need to better explain the exception/better document the investigation/better conclude on exception.</strong></td>
<td></td>
</tr>
<tr>
<td>Having the technical skills to write clear conclusions is another way to stand out at work.</td>
<td></td>
</tr>
<tr>
<td><strong>Again, you’ll need to work on that.</strong></td>
<td></td>
</tr>
</tbody>
</table>

“I know you haven’t finished yet, but I do need you to switch to another task now. You’ll get some written review notes on this AR task later.”

*Remainder of script transitions student to Inventory Task.*
Appendix C (Experiment 2)

Inventory Observations

The dependent variable in Experiment 2 is skeptical action, operationalized by measuring the participant’s choice to inquire further about an observation, and the participant’s choice to request supporting documentation based on an observation. The initial observation (after which the participant chooses whether or not to inquire) and the warehouse manager’s response (after which the participant chooses whether or not to request supporting documentation) are shown below. The explanations were provided in the form of an audio recording, coupled with an image of the warehouse manager with a speech bubble containing the full text of the response.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>While moving about the warehouse, you observe certain inventory that is segregated from the rest of the population, tagged with red tags instead of the usual green tag. This inventory is labeled SharpAct. These bins are not listed on your inventory listing.</td>
<td>These are the SharpAct items discussed in the count instructions I gave you earlier. We keep certain goods on consignment - they are owned by SharpAct. We still count them to report quantities to SharpAct, but we use red tags to keep the counts separate. These are our only consigned goods. Sam is well aware of this arrangement - nothing has changed since the last time I told him about it.</td>
</tr>
<tr>
<td>While moving about the warehouse, you observe one bin that appears old and dusty.</td>
<td>Sure, we have some slow-moving items in the warehouse. Sam knows all about it - it is typically clothing of non-standard sizes. Our accounting department checks up on these things regularly. You note the SKU number of the dusty box. You also obtain a copy of the last &quot;slow moving inventory report&quot; Graham prepared for the accounting department.</td>
</tr>
<tr>
<td>While moving about the warehouse, you observe a piece of equipment with a tarp over it and a sign that says &quot;out of service&quot;.</td>
<td>Yes, this piece is out of service. It’s been out about two months. We haven’t gotten any orders in that require it lately, so it hasn’t been high priority for maintenance. We should be getting it fixed soon. Listen - we really need to move things along. Sam and I agreed that we would be done at 9:30am, and I’m worried we’re going to go over.</td>
</tr>
<tr>
<td>Through general discussions you’ve heard among the workers, it sounds like there was a safety incident in the plant earlier this week.</td>
<td>Yes, that was pretty serious. There was an accident with one of our forklifts, and an employee was hospitalized. The company is in the process of doing a safety inquiry, but meanwhile they’ve sent out reminders regarding safety protocols. So Sam told me we’d be done by 9:30. I’ve got my team coming in to work at that time. You really need to get moving here.</td>
</tr>
<tr>
<td>While moving about the warehouse, you observe certain damaged items that had been set aside and excluded from the count.</td>
<td>This was in those instructions I gave you. If we notice a damaged good during the count, we set those goods aside and exclude them from the count. We also keep a damaged goods listing. Sam told me we’d be done by 9:30. Every minute we go over that time costs the company money both in wages and lost productivity. You need to get moving.</td>
</tr>
</tbody>
</table>
Both analyses resulted in bootstrap confidence intervals of the indirect effect that are significant at 99% Confidence Level.
Table 1
(Experiment 1)
Consideration of All Action Choices, Including Underreporting Time or Phantom Ticking

Panel A: Jack’s Anticipated Performance Rating by Condition and Action

<table>
<thead>
<tr>
<th>Condition</th>
<th>No proactive action</th>
<th>Proactive notification of overage</th>
<th>Under-report hours to meet budget</th>
<th>Skip steps to meet budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Reputation (PR) (n = 33)</td>
<td>5.15</td>
<td>7.91***</td>
<td>8.36***</td>
<td>5.09</td>
</tr>
<tr>
<td>Negative Reputation (NR) (n = 35)</td>
<td>3.66</td>
<td>6.08</td>
<td>7.46***</td>
<td>4.06</td>
</tr>
<tr>
<td>Negative false Reputation (NFR) (n = 30)</td>
<td>3.17</td>
<td>5.47</td>
<td>7.77***</td>
<td>3.87</td>
</tr>
</tbody>
</table>

Panel B: Significance Tests

Performance Rating: No action in response to anticipated overage

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>4.33</td>
<td>&lt;0.001a</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>4.88</td>
<td>&lt;0.001a</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>1.04</td>
<td>0.304</td>
</tr>
</tbody>
</table>

Performance Rating: Proactive notification of overage

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>3.82</td>
<td>&lt;0.001a</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>4.92</td>
<td>&lt;0.001a</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>1.07</td>
<td>0.291</td>
</tr>
</tbody>
</table>

Performance Rating: Under-report hours to meet budget

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>1.50</td>
<td>0.069a</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>0.92</td>
<td>0.180a</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>0.47</td>
<td>0.639</td>
</tr>
</tbody>
</table>

Performance Rating: Skip steps to meet budget

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>1.57</td>
<td>0.061a</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>1.86</td>
<td>0.034a</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>0.27</td>
<td>0.790</td>
</tr>
</tbody>
</table>

a These are directional tests, so p-values are one-tailed.
*** Significantly greater than 6 (scale midpoint, anchor of “acceptable”) p < 0.001.
Table 2
(Experiment 1)
Likelihood of Audit Staff Actions

Panel A: Likelihood Jack Will Choose Each Action

<table>
<thead>
<tr>
<th>Condition</th>
<th>Report all hours with proactive notification of overage</th>
<th>Report all hours with no proactive action</th>
<th>Under-report hours to meet budget</th>
<th>Skip steps to meet budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Reputation (PR) (n = 33)</td>
<td>38.33</td>
<td>19.70</td>
<td>28.27</td>
<td>13.70</td>
</tr>
<tr>
<td>Negative Reputation (NR) (n=35)</td>
<td>24.43</td>
<td>18.83</td>
<td>36.28</td>
<td>20.46</td>
</tr>
<tr>
<td>Negative false Reputation (NFR) (n=30)</td>
<td>31.33</td>
<td>16.47</td>
<td>37.47</td>
<td>14.73</td>
</tr>
</tbody>
</table>

Panel B: Significance Tests

Dependent Variable: Likelihood Jack will report all hours with proactive notification of overage

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>3.32</td>
<td>&lt;0.001^b</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>1.36</td>
<td>0.090^b</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>1.50</td>
<td>0.139</td>
</tr>
</tbody>
</table>

Dependent Variable: Likelihood Jack will report all hours (with or without proactive action)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>3.13</td>
<td>0.001^b</td>
</tr>
<tr>
<td>PR &gt; NFR</td>
<td>1.83</td>
<td>0.036^b</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>0.91</td>
<td>0.369</td>
</tr>
</tbody>
</table>

Dependent Variable: Likelihood Jack will under-report to meet budget

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &lt; NR</td>
<td>2.15</td>
<td>0.018^b</td>
</tr>
<tr>
<td>PR &lt; NFR</td>
<td>2.24</td>
<td>0.014^b</td>
</tr>
<tr>
<td>NR ≠ NFR</td>
<td>0.30</td>
<td>0.763</td>
</tr>
</tbody>
</table>

^a Participants allocated 100% over four possible actions: Report all hours with no proactive action; Report all hours with proactive notification of overage; Under-reporting hours to meet budget; Skipping steps to meet budget.

^b These are directional tests, so p-values are one-tailed.
Table 3  
(Experiment 2)  
Information Requests

Panel A: Descriptive Statistics of Information Requests

<table>
<thead>
<tr>
<th>Condition</th>
<th>Inquire Further</th>
<th>Request Documentation</th>
<th>Total Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td>Positive Reputation (PR)</td>
<td>3.00 (1.255)</td>
<td>1.60 (1.429)</td>
<td>4.60 (2.458)</td>
</tr>
<tr>
<td>(n=20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Reputation (NR)</td>
<td>2.33 (0.816)</td>
<td>0.80 (0.775)</td>
<td>3.13 (1.125)</td>
</tr>
<tr>
<td>(n=15)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Significance Tests

Dependent Variable: Inquire Further

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>1.79</td>
<td>0.042 a</td>
</tr>
</tbody>
</table>

Dependent Variable: Request Documentation

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>1.96</td>
<td>0.0294 a</td>
</tr>
</tbody>
</table>

Dependent Variable: Total Requests b

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR &gt; NR</td>
<td>2.14</td>
<td>0.020 a</td>
</tr>
</tbody>
</table>

a These are directional tests, so p-values are one-tailed.  
b While t-tests provide preliminary evidence, the level of documentation requests are dependent on whether the participant inquired further, and so we also conduct a Poisson regression to better fit our data. We first created a new dependent variable, skepticism, for each observation. Skepticism was coded as 0 if the participant neither inquired further nor requested documentation, 1 if the participant inquired further but did not request documentation, and 2 if the participant both inquired further and requested documentation. This resulted in 5 skepticism measures for each participant: one for each observation. We conducted a repeated measures Poisson regression to model the level of skepticism applied based on reputation condition, clustered by participant. Consistent with the main results, we find that the level of skepticism is significantly higher for participants in the positive reputation group (Z = 3.0, p = 0.003).