



Why Are People Honest? Internal and External Motivations to Report Honestly*

PAMELA R. MURPHY, *Queen's University*[†]

MICHAEL WYNES, *Wilfred Laurier University*

TILL-ARNE HAHN, *HEC Montréal*

PATRICIA G. DEVINE, *University of Wisconsin – Madison*

ABSTRACT

We create and validate measures capturing internal and external motivations to report honestly as trait-like characteristics. Both measures have high levels of reliability, as well as convergent and divergent validity. To test their predictive validity, we conduct two experiments. In the first experiment, MTurk participants have the opportunity and incentive to misreport with no immediate consequences, and in the second experiment, participants with management experience report how they would make a hypothetical accounting allocation decision. In both experiments, we find that participants who are higher in internal motivations to report honestly are more likely to report honestly than those lower in internal motivations, confirming this measure's predictive validity. Both experiments also provide limited support for the predictive validity of our external measure, finding that those who are higher in external motivation do not report differently than those who are lower in external motivations in the absence of controls. Our study also reveals that individuals who are higher in internal motivations have a diminished reaction to different management controls: MTurk participants to a control that punishes misreporting, and manager participants to a control that rewards honest reporting. Results suggest that management and those charged with governance should consider that some employees can react negatively to controls that are perceived as constraining. Our measures are useful to researchers who investigate honest reporting by allowing them to identify, *ex ante*, individuals who want to *be* honest versus wanting to *appear* honest.

Pourquoi les gens sont-ils honnêtes? Motivations intrinsèques et extrinsèques à communiquer honnêtement l'information

RÉSUMÉ

Les auteurs créent et valident des indicateurs permettant de saisir les motivations intrinsèques et extrinsèques, assimilables à des traits de caractère, à communiquer honnêtement l'information. Les deux types d'indicateurs affichent des taux élevés de fiabilité, ainsi qu'une validité convergente et divergente. Afin de tester la validité prédictive de ces indicateurs, les auteurs procèdent à deux

* Accepted by Jeffrey Cohen. The authors wish to thank Jeremy Bentley, Laurence Daoust, Michele Frank, Pujawati (Estha) Gondowijoyo, Jessica Osgood, Steve Salterio, Chez Sealy, and others from workshops at the University of Exeter, the University of Massachusetts–Amherst, Queen's University, and the AAA Forensic Accounting section meeting and annual meeting for their helpful comments on an earlier version. We thank Qiu Chen, Khim Kelly, and Steve Salterio, as well as Jeff Cohen, Laurie Pant and David Sharp, for the use of their case. We also thank the two anonymous reviewers and Jeff Cohen for their insightful comments. Pam thanks the Smith School of Business for its continuing financial support.

[†] Corresponding author.

expériences. Dans la première, des participants travailleurs du service MTurk ont l'occasion de communiquer de l'information inexacte et sont incités à le faire, sans conséquences immédiates. Dans la seconde, des participants expérimentés en gestion font état de la façon dont ils prendraient une décision de répartition comptable hypothétique. Dans l'une comme dans l'autre de ces expériences, les auteurs constatent que les participants dont les motivations intrinsèques à communiquer honnêtement l'information sont plus élevées sont davantage susceptibles de le faire que les participants dont les motivations intrinsèques sont plus faibles, ce qui confirme la validité prédictive de cet indicateur. Par ailleurs, les deux expériences appuient modérément la validité prédictive de leur indicateur extrinsèque, démontrant que les participants dont les motivations extrinsèques sont plus élevées ne communiqueront pas l'information de façon différente de ceux dont les motivations extrinsèques sont plus faibles, en l'absence de contrôles. L'étude révèle également que les participants dont les motivations intrinsèques sont plus élevées ont une réaction moins marquée aux différents contrôles de gestion : les participants du service MTurk exposés à un contrôle en vertu duquel la communication d'information inexacte est sanctionnée, et les participants gestionnaires exposés à un contrôle en vertu duquel l'honnêteté dans la communication de l'information est récompensée. Les résultats laissent croire que les dirigeants et les personnes à qui incombe la gouvernance devraient tenir compte du fait que certains employés peuvent réagir négativement aux contrôles perçus comme étant contraignants. Les indicateurs proposés par les auteurs seront utiles aux chercheurs qui s'intéressent à l'honnêteté dans la communication de l'information, en leur permettant de distinguer, *ex ante*, les personnes qui désirent être honnêtes de celles qui souhaitent paraître honnêtes.

1. Introduction

The conundrum of why individuals report honestly continues to interest researchers. Those steeped in economics attempt to understand why so many people are honest even if they have no apparent reason to be, whereas those steeped in psychology wonder why so many people misreport when they should not (Salterio and Webb 2006). In the accounting literature, numerous studies examine (dis)honest reporting, with many recognizing that individuals vary with regard to preferences for honesty (e.g., Church et al. 2014; Evans et al. 2001; Luft 1997; Rankin et al. 2008) and others acknowledging that individuals often prefer to appear honest (Hannan et al. 2006). We argue that wanting to *be* honest is different from wanting to *appear* honest. According to self-determination theory, people who want to *be* honest should report more honestly across all circumstances, because of their internal motivation; however, people who want to *appear* honest will report more honestly only when they believe their behavior is observed (Deci et al. 2017; Ryan and Deci 2000). The overarching objective of this paper is to create and validate two measures: (i) internal motivations to report honestly, which capture the desire to *be* honest, and (ii) external motivations to report honestly, which capture the desire to *appear* honest. Additionally, we examine whether individuals higher in internal motivations react differently to management controls that encourage honest reporting versus those that punish misreporting.

It is important to understand motivations to report honestly for several reasons. First, motivations to report honestly are not as well understood as those to misreport. Prior research has validated several motivations to misreport, such as financial incentives or pressures (Hogan et al. 2008; Trompeter et al. 2013) or social incentives or pressures (Church et al. 2012; Mayhew and Murphy 2014), and these motivations are recognized in audit standards (IAASB 2009; PCAOB 2005), as well as in forensic accounting practice (Dorminey et al. 2011). However, the motivation to report honestly often emanates from within an individual, such as one's own beliefs and ethical values. Other research has recognized that individuals are driven in part by noneconomic preferences (Gibbons 1998). Psychology, in particular, has examined moral identity as a character trait (Aquino and Reed 2002; Miles and Upenieks 2018), moral judgments that apply intuition from social and cultural cues, as well as rationality (Haidt 2001) and moral foundations (Graham et al. 2011; Haidt et al. 2009). However, the ability to specifically identify and measure honest reporting as an individual motivation remains unresolved.

Second, although dishonest reporting has been extensively studied,¹ much of the literature appears to assume that people will react the same way to controls designed to prevent dishonesty versus controls that promote honesty. We argue this may not be the case. According to Salterio and Webb (2006) and organizational behavior researchers (Grover 2005; Murphy 1993), promoting honesty and preventing dishonesty in the workplace are two different things. Some individuals can react negatively to certain management controls, resulting in behavior counter to what is intended. Self-determination theory and motivation crowding theory both suggest that those who are higher in internal motivations to report honestly may rebel against a control that is perceived to be controlling or threatening, leading them to actually *misreport* (Deci and Ryan 2012; Frey and Jegen 2001). Finally, we are unaware of any research that has decomposed internal and external motivations to report honestly as an individual trait.

We follow the process put forth by Churchill (1979) and elaborated by Sandino (2011) to create our measures. After developing items for our measures, we test and retest them, and examine convergent and divergent validity. Our two measures—internal and external motivations to report honestly—represent unique, uncorrelated traits with strong coefficient alphas, and they correlate as expected with other measures designed to capture internal and external motivations in other domains.

We test our measures' predictive validity using two complementary experiments. First, we allow MTurk participants to make real reporting decisions—we pay them the amount they report—and we offer salient potential rewards and punishments to encourage honest reporting. Second, we ask Prolific participants with management experience to make a hypothetical accounting decision and we simulate similar rewards and punishments. Even though we expect the same results, we add to the validity of our findings by conducting two experiments in different settings. In both, we find that participants higher in internal motivations to report honestly are, in fact, more likely to report honestly than those lower in internal motivations. Our internal motivation measure is valid and predicts behavior. We find limited support for our external measure. That is, while those higher in external motivations do not report any differently than those lower in external motivations when management controls are not present, as expected, we do not find that those higher in external motivations actually report more honestly when management controls are present.

Finally, for individuals higher in internal motivations, we observe that certain management controls have adverse effects. In the first experiment, we find that these individuals' reactions to management controls that punish misreporting are attenuated. Specifically, the propensity to report more honestly in response to this condition is *less* pronounced for those who are higher in internal motivations. We find a similar effect in the second experiment, except that the response is to management controls that reward honest reporting. These reactions are consistent with the underlying theories, because the participants also report feeling constrained.

Our findings have potential implications for management and those charged with governance: They should be aware of their employees' individual motivations and attempt to adapt the types of management controls to them. Failing to take this into account may lead to unintended consequences. For example, if management determines that most employees in a particular department are higher in internal motivations, it should ensure that the controls and monitoring of those individuals are not perceived as too constraining or they may backfire.

Overall, we extend the understanding of why people report honestly in circumstances in which such behavior may not be expected or cannot be accurately predicted. For example, in an experiment in which participants were able to lie with no consequences, Gibson et al. (2013) observe heterogeneity in individuals' preferences for truthfulness, with some on either end of the spectrum and with many—those whose reporting behavior shifts as the costs of truthfulness increase—in the middle. Our measures are able to capture some of this heterogeneity. In an

1. In addition to the accounting literature, researchers in other fields study honesty, such as Ariely (2012) in psychology and Gneezy et al. (2013) in behavioral economics.

accounting study examining the impact of information acquisition discretion on opportunistic reporting, Church et al. (2014) first conducted a series of baseline conditions designed to identify individuals having high versus moderate versus low honesty preferences. Our internal measure can identify individuals' preferences for honesty without having to conduct preliminary experiments. Relatedly, Hannan et al. (2006) argue that individuals have preferences for appearing honest, and Hao and Houser (2017) find that required announcements of future behavior result in more honest behavior by those who care about maintaining "an honest self-image." Our external measure, having convergent and divergent validity, can identify individuals' preferences for appearing honest. In essence, our measures can be used in future work on honesty in order to distinguish these individual motivations, *ex ante*.²

This paper also adds to the auditing literature. For example, synthesizing the management and internal control literature, Fiolleau et al. (2018) develop audit research ideas. Our measures address their call for more study into the role of individual differences in assessing fraud risk, because fraud risk could be assessed as lower for individuals who are higher in internal motivations to report honestly.³ Our measures also indirectly assist with some of their other research ideas in the following ways: (i) the finding that high monitoring may have a deleterious impact on a workforce that is high in internal motivations should increase inherent risk assessments; (ii) knowing that private information held by subordinates may not be revealed in organizations should impact audit risk assessment, but our measures suggest that this may not be as problematic with employees who are higher in internal motivations; and (iii) management integrity assessments might be aided by our measures.

Finally, we believe that our measures have relevance to practice. Our measures could be incorporated into personality tests for hiring purposes.⁴ Perhaps the best form of contracting with an employee who is high in internal motivations to report honestly is a trust contract (Church et al. 2014), because such individuals are more likely to report honestly without the need for external reinforcement. Additionally, it would be helpful to incorporate these measures, as well as questions soliciting employees' perceptions of management controls, into annual employee surveys. Administering regular employee surveys is a recognized high-performance human resource practice (Judd et al. 2018). If managers know the motivational tendencies of a group of employees, they can consider those predispositions when implementing controls. For example, if a certain group of employees is higher in internal motivations to report honestly, management should also inquire as to employees' perceptions of the existing management controls. If these employees perceive the controls as threatening, constraining, or controlling, management should consider replacing those controls with softer, more supportive controls.

2. Development and validation of the measures

We begin with a definition of the constructs we wish to capture. We are interested in enhancing our understanding of why individuals report honestly when they have no perceived need to do so (e.g., no consequences), and we believe that such motivations, at least in part, can be represented and captured as trait-like characteristics. We define honest reporting as reporting the truth or reality to others, whereas misreporting is not. For example, individuals may misreport the income on their tax returns, employees may misreport their hours worked, or managers may misreport their division's or organization's income. We believe that motivations to report honestly can be

-
2. Our measures are reported in Table 4, and the entire instrument is available in the online supplement. Please see supporting information, as an addition to the online article.
 3. Fraud risk is lower if top management consists mostly of individuals higher in internal motivations, because frauds committed by high-level perpetrators are more costly than those committed by lower-level employees (ACFE 2018).
 4. Basham (2011) reports that some banks use psychometric testing to recruit people with psychopathic tendencies for senior corporate positions. Whether or not they are used directly as a recruitment instrument, we suggest our measures could be used to build awareness of the types of people being hired.

characterized as internal or external. We partially follow work that has been done on similar constructs in psychology,⁵ using self-determination theory as our primary theoretical foundation.⁶

Self-determination theory (Ryan and Deci 2000) posits that there are various motivations to behave in a certain way. The theory has identified specific types of motivations, ranging from completely external to completely internal. If a motivation is external (extrinsic), the individual performs the act for instrumental purposes (e.g., to achieve a specific outcome). If internal (intrinsic), the person performs the act for its own sake (i.e., for the inherent satisfaction of having performed it).⁷ Beginning with the extreme end of external motivations, individuals act in a certain way purely to earn external rewards or to avoid external punishments. As individuals begin to exercise self-control, they still have external motivations, but those motivations become more internally oriented (e.g., to avoid guilt or to build their ego). This type of external motivation has to do with internal rewards or punishments. Toward the middle of the continuum, individuals appear to have both external and internal motivations. They consciously value the act and find it to be personally important; however, they do not fully find it intrinsically satisfying. The extreme internal end is where individuals perform the act purely for the intrinsic satisfaction of having performed it or the inherent belief in the act itself. Although we are not interested in parsing out all of these degrees, we wish to decompose motivations to report honestly into primarily internal or external.⁸

Developing the items

We began generating sample items by examining answers to a question posed in prior research for which no analysis had previously been performed (Mayhew and Murphy 2014; Murphy 2012).⁹ In both experiments, participants completed a task that generated an earned income; they were then able to report any income (within the parameters of the income range) and were paid the amount they reported. Upon finishing, participants were asked to explain why they reported as they did. We examined and coded the answers for those who reported honestly.¹⁰ Table 1 presents the resulting categories.

The six groups capture a mix of motivations. One group, being satisfied with their earned income (10 percent of the total), appears consistent with economic theory. Another, avoiding hurting someone (19 percent), is consistent with literature on how concerns about fairness and equity influence reporting decisions (Cohen et al. 2007; Cohen et al. 2015; Drake et al. 2014; Fehr and Schmidt 1999; Maas and Van Rinsum 2013). The desire not to hurt someone else can also be linked to the notion of caring, which, similar to fairness, has been associated with

-
5. For example, Plant and Devine (1998) derive internal versus external motivations to respond without prejudice. This seminal paper led to many studies, including research on interventions to reduce racial bias (Devine et al. 2012).
 6. Theories from different fields may apply to this construct. For example, conventional economic theory suggests that individuals will misreport when given the opportunity to misreport without consequence. Consequences can extend beyond the economic: they can also be social. We capture social consequences in our measure of external motivations to report honestly. Additionally, economic studies have shown that many people still report honestly (Gneezy et al. 2013; López-Pérez and Spiegelman 2013) when there are no consequences. Although an individual's own utility function will capture an internal desire to be honest, research has not measured it *ex ante*. We capture this construct with our measure of internal motivations to report honestly.
 7. Ryan and Deci (2000) refer to motivations as extrinsic and intrinsic. We refer to them as external and internal.
 8. Internal motivations depend less upon contextual factors, whereas external motivations, by their nature, are context dependent. For example, the moral intensity of the issue itself (Jones 1991) might impact how someone with higher external motivations reacts. Specifically, the higher the moral intensity, the more likely the person will report honestly.
 9. With permission from the authors.
 10. Two authors independently coded the answers. Inter-rater agreement was 92 percent (Cohen's kappa = 0.89, $p < 0.01$) and 81 percent (Cohen's kappa = 0.68, $p < 0.01$), respectively, from the Murphy (2012) and Mayhew and Murphy (2014) data. After further discussion, all discrepancies were resolved.

TABLE 1

Reasons for reporting honestly (from Murphy (2012) and Mayhew and Murphy (2014))

Categorization	Number from Murphy (%)	Number from Mayhew and Murphy (%)	Total (%)	Examples
Self-worth/dignity	1 (1)	3 (6)	4 (2)	“It is not worth my dignity” and “I have my integrity”
Fear of getting caught	11 (7)	6 (11)	17 (8)	“I was afraid to be the only one to submit a fake number” and “you never know who may find out your actions”
Satisfied with earned income / amount not enough to bother with	17 (11)	3 (6)	20 (10)	“I think \$19.87 is a reasonable amount, I might have reported more if say I got \$5” and “the few dollars aren’t enough incentive for me to cheat or lie”
Avoiding negative emotion	14 (9)	8 (15)	22 (11)	“I just knew I would feel guilty the minute I walked out the door” and “I thought I would eventually regret it”
Avoiding hurting someone	39 (25)	0 (0)	39 (19)	“I would be cheating someone else for my own personal gain” and “my reported answers [affects] another student”
Ethics	71 (46)	34 (63)	105 (51)	“I wanted to make an ethical decision,” “moral values,” “it would be against my principles,” and “it did not seem like the right thing to do”
Totals ^a	153	54	207	

Notes: Values in parentheses are indicated in percentage and may not add to 100 due to rounding. Murphy (2012) used a manipulation in which any misreporting by a participant would hurt another participant by reducing his/her payment. This accounts for the large number of honest reporters indicating they wanted to avoid hurting someone. Mayhew and Murphy (2014) used a manipulation in which the person administering the experiment (the “boss”) asked participants to misreport. ^aThe original articles report 147 (71 percent) honest reporters (Murphy 2012) and 57 (48 percent) honest reporters (Mayhew and Murphy 2014). The discrepancy is because we coded all useable responses, whereas the previous articles excluded individuals who had missing data from other parts of the experimental materials (e.g., affect measures in Murphy (2012)) or included individuals who did not provide written responses to the open-ended question about honest reporting (Mayhew and Murphy 2014).

morality (Graham et al. 2013).¹¹ The fear of getting caught (8 percent) represents an external motivation, consistent with a pre-conventional level of ethical reasoning, in which individuals decide right and wrong based upon the anticipated consequences (Jones et al. 2003; Kohlberg 1969). Finally, three groups appear to capture internal motivations to report honestly: self-worth/dignity (2 percent), avoiding negative emotion (11 percent) and the largest category, ethics (51 percent), which appears to capture answers relating to morals, principles, and general feelings of right and wrong.

11. Wishing to avoid hurting someone else can be explained by the manipulations in that particular study (Murphy 2012), in which any misreporting financially harmed a specific participant who was on student financial aid.

Following self-determination theory, we then classified the categories as external or internal. Ryan and Deci (2000) describe three categories of motivation that are primarily external: (i) external regulation, in which external rewards or punishments dictate behavior, (ii) introjected regulation, in which the motivation has a bit more of an internal element such as internal rewards (e.g., enjoying one's good reputation) or punishments (e.g., suffering one's poor reputation), and (iii) identified regulation, in which the motivation exists because the behavior is required, even though the person does not accept it internally. For example, "I report honestly to get approval" represents an external regulation, "Reporting honestly helps my reputation" exemplifies an introjected regulation, and "I avoid misreporting because it is against the rules" is an identified regulation.

The two categories of internal motivation from Ryan and Deci (2000) include (i) integrated regulation, in which the behavior is important to the individual because it is fully assimilated into oneself, and (ii) intrinsic regulation, in which the person performs the behavior for the sheer value, interest, or inherent satisfaction it provides. For example, "Reporting honestly makes me feel good about myself" captures an integrated regulation, whereas "Misreporting is against my personal values" represents an intrinsic regulation.

We brainstormed other items that we believed represented external and internal motivations, using theory and feedback from colleagues. Our aim was not to categorize every item into Ryan and Deci's (2000) framework or prior literature but rather to ensure completeness. Our final list of statements after preliminary testing—12 external and 12 internal motivation items—is reported in Table 2.¹²

Validating the measures

We administered our measures using an online survey with MTurk workers.¹³ Participants were paid US\$3.75 each for their participation. Our sample consisted of 604 participants with an average age of 35 years. Fifty-four percent were male, 53 percent college-educated or higher, and 66 percent employed full-time.¹⁴ See Table 3 for the demographics of Sample 1.

We first performed correlation analyses (not reported herein) of the items theorized to be associated within each dimension. The highest Pearson correlations are among our internal items, all of which are significant (two-tailed) and with the highest correlation being 0.83. The correlations among our external items are more variable, ranging from nonsignificant to the highest being 0.67.

Next, we performed a principal component factor analysis with oblimin rotation (Nunnally and Bernstein 1994; Plant and Devine 1998) on all 24 items. The first outcome revealed three factors, one accounting for 41 percent of the variance, the second, 17 percent, and the third, 5 percent (all with eigenvalues over 1.00). We then used an iterative process in which we removed any variable that cross-loaded across factors, one at a time. When no variables cross-loaded, we removed the weakest loading variable, one at a time, while ensuring that the overall variance explained by the results did not drop significantly. We retained variables that loaded at 0.70 or greater. The solution of this process yields two factors that, together, account for 72 percent of the variance, the first for 51 percent and the second, 21 percent. In Table 2, we report all the factor loadings and bold the items retained after performing our factor analysis. Four items capture a theme of wishing to please others as the external motivation factor, with a Cronbach's alpha of 0.84. Nine items capture the internal motivation factor, with a Cronbach's alpha of 0.95 and a theme

12. We performed preliminary testing with an earlier list of 20 items (not reported herein), surveying both students ($n = 551$) and MTurk workers ($n = 336$). We also attempted to delineate motivations to report honestly versus motivations to avoid misreporting. Correlation and factor analyses yielded no such distinction in the data.

13. Before conducting any survey or experiment, we secured ethics approval. All participants provided their informed consent prior to participating.

14. As labor force participation rates are typically in this range, we consider our sample to be representative of the overall adult population. Nevertheless, for robustness we reran our analysis of the scale development (Sample 1) and Experiment 1 (Sample 2) results after dropping MTurk participants who (i) did not have at least part-time employment and (ii) had not taken some university level courses. We discuss this further in the appropriate sections.

TABLE 2
Complete list of items for external and internal motivations to report honestly

Items	Initial factor loadings	Final factor loadings
External motivations to report honestly or avoid misreporting		
E1: It is important to me that other people think I report honestly	0.35	
E2: If I report honestly, others will respect me more	0.38	
E3: I report honestly if it pleases others	0.82	0.84
E4: Reporting honestly helps my reputation	0.31	
E5: I report honestly to get approval	0.80	0.83
E6: I report honestly if I think I will receive acknowledgement for doing so	0.80	0.84
E7: If I misreported, I would be concerned about losing the respect of people important to me	0.17	
E8: I would be concerned about my reputation if I misreported	0.22	
E9: Others would disapprove if I misreported	0.14	
E10: I would avoid misreporting if it pleased others	0.78	0.79
E11: I avoid misreporting because it is against the rules	0.20	
E12: The benefit of misreporting is not worth the risk of being punished	0.06	
Internal motivations to report honestly or avoid misreporting		
I1: Reporting honestly makes me feel good about myself	0.71	0.77
I2: Reporting honestly feels like the right thing to do	0.82	0.84
I3: Reporting honestly is important to me	0.91	0.92
I4: I value honest reporting for its own sake	0.87	0.90
I5: Reporting honestly maintains my self-worth	0.74	0.78
I6: I believe in reporting honestly even if no one will know	0.90	0.90
I7: Misreporting is against my personal values	0.88	0.87
I8: I would feel guilty if I misreported	0.81	0.81
I9: Misreporting does not feel like the right thing to do	0.67	
I10: Misreporting is not consistent with the person I see myself being	0.74	
I11: I would have trouble sleeping if I misreported, even if there were no chance of being caught	0.63	
I12: No matter what others think or believe, it is important not to misreport	0.89	0.89

Notes: These were administered using a nine-point Likert scale from 1 = “strongly disagree” to 9 = “strongly agree.” The bolded items are the statements retained after performing factor analysis on Sample 1. Four items capture external motivation (E3, E5, E6, and E10), and nine capture internal motivation (I1 through I8 and I12). Initial factor loadings of each item are reported first, as well as the final factor loadings for the retained items.

capturing internal values, avoidance of negative feelings associated with misreporting, and the general importance of reporting honestly. In summary, our measures have high Cronbach’s alphas, exceeding what is considered very good, and they capture identifiable themes.¹⁵

To further validate our measures, we collected a second sample, which we also administered through MTurk. Participants were paid US\$2.00 for completing the survey. Table 3 reports the demographics of the full Sample 2 ($n = 681$) and a reduced Sample 2 ($n = 600$).¹⁶ We compared

15. DeVellis (1991) and Hurtt (2010) indicate that a Cronbach’s alpha of 0.65 to 0.70 is minimally acceptable, a range between 0.70 and 0.80 is respectable, while 0.80 and 0.90 is considered very good.

16. Because we administered our survey and first experiment using Sample 2, we reduced the sample size for participants who failed our comprehension and manipulation check questions. We discuss this in greater depth, along with participant payments, in section 3.

TABLE 3
Demographics of Samples 1, 2, and 3

Item	Sample 1 (<i>n</i> = 604)	Full Sample 2 (<i>n</i> = 681)	Reduced Sample 2 (<i>n</i> = 600)	Full Sample 3 (<i>n</i> = 224)	Reduced Sample 3 (<i>n</i> = 126)
Average age	35	37	37	37	39
Youngest	18	20	20	19	19
Oldest	72	84	84	67	67
Gender					
Male	327 (54)	348 (51)	309 (51)	102 (46)	52 (41)
Female	274 (45)	331 (49)	289 (48)	121 (54)	73 (58)
No answer	3 (0.5)	2 (0.2)	2 (1)	1 (0.3)	1 (1)
Education					
High school	95 (16)	94 (14)	79 (13)	9 (4)	6 (5)
Some university	188 (31)	236 (35)	208 (35)	52 (23)	28 (22)
Completed university	261 (43)	277 (41)	245 (41)	85 (38)	51 (40)
Completed Masters	53 (9)	53 (8)	48 (8)	49 (22)	27 (21)
Other	7 (1)	21 (3)	20 (3)	28 (12)	14 (11)
No response	0 (0)	0 (0)	0 (0)	1 (0.3)	0 (0)
Current student status					
Full-time student	75 (12)	43 (6)	39 (7)	n/a	n/a
Part-time student	28 (5)	18 (3)	15 (3)	n/a	n/a
Not a student	501 (83)	620 (91)	546 (91)	n/a	n/a
Current employment status					
Full-time employed	399 (66)	434 (64)	377 (63)	184 (82)	103 (82)
Part-time employed	103 (17)	121 (18)	113 (19)	36 (16)	21 (17)
Not employed	102 (17)	126 (19)	110 (18)	4 (2)	2 (2)
Work experience					
None	16 (3)	22 (3)	21 (4)	1 (0.5)	1 (1)
Less than one year	18 (3)	9 (1)	7 (1)	0 (0)	0 (0)
1 to 5 years	114 (19)	116 (17)	108 (18)	34 (15)	10 (8)
6 to 10 years	142 (24)	152 (22)	136 (23)	41 (18)	24 (19)
11 to 20 years	181 (30)	195 (29)	173 (29)	88 (39)	52 (41)
More than 20 years	133 (22)	187 (27)	155 (26)	60 (27)	39 (31)
Average management experience	n/a	n/a	n/a	7	8
Minimum	n/a	n/a	n/a	0.5	1
Maximum	n/a	n/a	n/a	48	48
Country of birth					
United States	584 (97)	654 (96)	578 (96)	92 (41)	53 (42)
United Kingdom	n/a	n/a	n/a	119 (53)	68 (54)
Canada	4 (1)	6 (1)	6 (1)	n/a	n/a
Other	16 (3)	21 (3)	16 (3)	13 (6)	5 (4)

Notes: Values in parentheses are indicated in percentage and may not add to 100 due to rounding.

participants' age, gender, education level, work status, and work experience between Sample 1 and Sample 2 using a chi-squared test. Results confirm the two samples are not statistically different in their demographic composition (all *p*-values > 0.05).

We performed confirmatory factor analyses with both the full and reduced Sample 2. Goodness-of-fit indices are identical for the full and reduced samples (CFI = 0.96, SRMR = 0.06, RMSEA = 0.08), with the exception of chi-square/*df*, which is 5.59 for the full sample and 4.83

for the reduced sample. This difference is driven by the number of observations in the data. Based on suggested levels for acceptable models, we conclude that the internal and external constructs demonstrate a moderately good fit to the data (Hu and Bentler 1999; Iacobucci 2010).¹⁷ Using a maximum likelihood model, the goodness-of-fit test is significant at $p < 0.01$ for both the full and the reduced samples. The Cronbach's alpha of our external factor is 0.82, while the Cronbach's alpha of our internal factor is 0.95 for both the full and the reduced samples. Similar to results from Sample 1, the consistently high Cronbach's alphas lend further credibility to the internal consistency of our measures.

Examining convergent and divergent validity

According to Churchill (1979), a good measure should correlate with other measures designed to capture similar constructs. Specifically, it should correlate highly with other similar measures, exhibiting convergent validity, and it should be novel and not just a reflection of another measure, displaying divergent validity. If evidence of these types of validity cannot be established, then a measure may not be capturing what it purports to represent.

We begin by arguing that our internal and external measures should not be correlated with each other, consistent with Plant and Devine (1998). Self-determination theory argues that internal and external motivations are distinct constructs because they emanate from different sources (from within and outside the individual respectively). Next, we used Plant and Devine (1998) as a guide to identify seven scales we expect to be correlated with high internal or high external motivations, which are as follows: (i) Fear of Negative Evaluation; (ii) Self-Monitoring; (iii) Honesty-Humility; (iv) External/Internal Self-concept; (v) Self-worth; (vi) Self-control; and (vii) the dark triad, consisting of narcissism, Machiavellianism, and psychopathy. Part B of the online supplement describes each of these measures in greater detail, while panel B of Table 4 reports our correlation expectations for each in the correspondingly labeled columns.

The Fear of Negative Evaluation scale (Leary 1983) is designed to assess an individual's fear of being negatively evaluated by others. We expect it to be positively associated with external motivations to report honestly because it captures concerns for external rewards and punishments. The Self-Monitoring scale (Snyder and Gangestad 1986) assesses an individual's self-observation and self-control guided by situational cues to social appropriateness. We expect this measure to be positively correlated with external motivations to report honestly. The Honesty-Humility domain of the HEXACO Personality Inventory (Lee and Ashton 2004) captures sincerity, fairness, greed avoidance, and modesty. We expect a positive relation between this scale and internal motivations to report honestly, because it captures personality traits associated with one's values. The Motivation Sources Inventory questions measure internal and external self-concept (Barbuto and Scholl 1998), capturing motivations that are both externally and internally reinforced. We expect a positive correlation between internal (external) self-concept and internal (external) motivations to report honestly. The Rosenberg Self-Esteem Inventory (Rosenberg 1965) measures self-worth. Given that some of the answers to why they reported honestly from Table 1 relate to self-worth and given that an item about self-worth is a part of our internal motivation factor, we expect a positive association between this scale and internal motivations to report honestly.

We also include a measure of self-control from the Brief Self-Control scale (Tangney et al. 2004), arguing those who are high in internal motivations to report honestly would need to exert some form of self-control in order to remain consistent with their beliefs. Thus, we expect a positive association between the self-control measures and internal motivations. Finally, we selected the short version of the dark triad of narcissism, Machiavellianism, and psychopathy,

17. Hu and Bentler (1999) suggest that good fitting models have CFI > 0.95, SRMR < 0.08, RMSEA < 0.06. Iacobucci (2010) notes that the raw Chi-square is impacted by large sample sizes and that the more appropriate measure is the Chi-square scaled by degrees of freedom. The recommended level for this measure is less than three.

TABLE 4
Results of analyses on convergent and divergent validity, using Sample 1

Panel A: Descriptive statistics					
Factor	Mean (SD)		Cronbach's alpha		
External (motivation for reporting honestly)	5.27 (2.05)		0.84		
Internal (motivation for reporting honestly)	7.83 (1.34)		0.95		
Fear of negative evaluation	2.91 (1.07)		0.96		
Self-monitor	-0.05 (0.24)		0.82		
Honesty-humility	6.02 (1.54)		0.89		
External self-concept	5.15 (1.65)		0.81		
Internal self-concept	7.22 (1.17)		0.81		
Self-worth	6.68 (1.96)		0.95		
Self-control	6.08 (1.58)		0.85		
Narcissism	3.61 (1.96)		0.84		
Machiavellianism	3.32 (1.93)		0.86		
Psychopathy	3.24 (1.71)		0.78		

Panel B: Expectations and correlations of external and internal motivations to report honestly with other measures					
Scale	Expected relation with external	Correlation with external	Expected relation with internal	Correlation with internal	Expectation supported?
External motivation factor			None	0.02	Yes
Internal motivation factor	None	0.02			Yes
Fear of negative evaluation	+	0.19		-0.05	Yes
Self-monitor	+	0.15		-0.25	Yes
Honesty-humility		-0.32	+	0.47	Yes
External self-concept	+	0.52		-0.13	Yes
Internal self-concept		0.05	+	0.47	Yes
Self-worth		-0.06	+	0.19	Yes
Self-control		-0.12	+	0.33	Yes
Narcissism		0.32	-	-0.31	Yes
Machiavellianism		0.19	-	-0.48	Yes
Psychopathy		0.12	-	-0.55	Yes

Notes: The External factor consists of the following four items. E3: I report honestly if it pleases others; E5: I report honestly to get approval; E6: I report honestly if I think I will receive acknowledgment for doing so; and E10: I would avoid misreporting if it pleased others. The Internal factor consists of the following nine items. I1: Reporting honestly makes me feel good about myself; I2: Reporting honestly feels like the right thing to do; I3: Reporting honestly is important to me; I4: I value honest reporting for its own sake; I5: Reporting honestly maintains my self-worth; I6: I believe in reporting honestly even if no one will know; I7: Misreporting is against my personal values; I8: I would feel guilty if I misreported; and I12: No matter what others think or believe, it is important not to misreport. Bolded items are significant at $p \leq 0.05$ two-tailed. For correlations without an expected relation (+ or -) we do not have any a priori expectations with our scale.

as these traits have been linked to misreporting and related unethical behavior (Paulus and Williams 2002). Because these traits essentially represent the opposite characteristics of internal motivations to report honestly, we expect each of them to be negatively associated with our internal factor.

We administered these measures to the same sample of MTurk participants described earlier (Sample 1). Panel A of Table 4 reports the mean (SD) of each factor along with its coefficient alpha. All factors have a Cronbach's alpha exceeding 0.70, which is considered respectable (DeVellis 1991; Hurtt 2010). Panel B of Table 4 presents the results of our analysis of the expected correlations. We note that our external and internal factors are not significantly correlated with each other (0.02), indicating that they represent two different constructs (consistent with self-determination theory) and suggesting that being low in one factor does not necessarily equate to being high in the other. As expected, there is a positive correlation between our external factor and fear of negative evaluation, self-monitoring, and external self-concept. Also as expected, there is a positive correlation between our internal factor and honesty-humility, internal self-concept, self-worth, and self-control. In addition, we noted that despite the external and internal factors not being correlated, being high (low) in one factor equates to being low (high) in the other, such as for the dark triad personality traits, external self-concept, honesty-humility, and self-monitoring. While a correlation was expected in each case with one of the factors (as outlined in Table 4, panel B), we did not expect the inverse correlation with the other factor.

We are intrigued by the positive correlation between the dark triad and the external factor. The three character traits of the dark triad have been linked to misreporting and unethical behavior (Ham et al. 2017; Majors 2016; Murphy 2012).¹⁸ Our correlation suggests that the motivation to appear honest is similar to traits associated with misreporting; it suggests that individuals who want to appear honest also have a predilection to misreport. The implication for management and those charged with governance could be severe, as it implies that employees who are higher in external motivations might not be trustworthy. We believe there is far more nuance to this correlation and future research should examine how these individual motivations interact in reporting settings.

Summary of section 2

To summarize this section, we developed items for both our measure of internal and external motivations, relying on prior research (Mayhew and Murphy 2014; Murphy 2012), theory (Ryan and Deci 2000), and feedback from colleagues, and by conducting extensive analysis with different participant samples. We identified four items that capture an external motivation to report honestly and nine items that capture an internal motivation. The Cronbach's alphas of both measures exceed 0.80, which is considered very good (DeVellis 1991; Hurtt 2010). Finally, we find evidence consistent with our expectations regarding measures that should and should not be correlated with ours, supporting their convergent and divergent validity. We are confident that we developed two measures that capture unique constructs, yet they are related to similar constructs as expected. We can eliminate the possibility that our measures are not simply an artifact of the way in which we created and analyzed the items.¹⁹

3. Examining predictive validity and the impact of controls

Even if a measure is internally consistent and has convergent and divergent validity, it cannot be considered reliable if it does not predict behavior. According to Churchill (1979), the final test of a good measure is to examine whether it behaves as expected. In this section, we test the predictive validity of our two measures and examine the interactions of our measures with two types of controls: (i) those that reward honest reporting and (ii) those that punish misreporting. We administered two experiments because we were unable to simultaneously use actual managers and

18. Interestingly, Lee et al. (2013) find the dark triad to be highly correlated with low honesty-humility, similar to our finding of a negative correlation between our external measure and honesty-humility.

19. For robustness we also ran our analysis on Sample 1 data after dropping MTurk participants who (i) did not have at least part-time employment and (ii) had not taken some university level courses. We find that excluding these participants has no effect on the scale items selected, nor does it significantly change results on the means, SD, or Cronbach's alphas of the other measures.

measure actual behaviors. Both experiments have three conditions: (i) a baseline that provides the opportunity and financial incentive to misreport, (ii) a condition that rewards honest reporting, and (iii) a condition that punishes misreporting. The first experiment was conducted via MTurk (Sample 2) with individuals from the general population and the second via Prolific (Sample 3) with individuals having management experience (see demographics for Samples 2 and 3 in Table 3). Although the two experiments are analogous in design, they test two distinct dependent variables: (i) real reporting behavior with MTurk participants who are paid the amount they report and (ii) a hypothetical accounting decision with manager participants. Next, we examine the accounting literature on misreporting and then develop our hypotheses to test our measures' predictive validity and two research questions to consider the impact of controls. Finally, we discuss each experiment and its results.

Accounting literature on misreporting

The management accounting literature has had a long history of examining (dis)honesty in reporting (see, e.g., Lowe and Shaw 1968 or Baiman and Lewis 1989). Much of it is in a budgeting context, using economic theory—predominantly agency theory—as the theoretical foundation. However, portraying a world of entirely self-interested individuals who are uninfluenced and unmotivated by anything other than simple cost-benefit calculations, agency theory has the problem of being a self-fulfilling prophecy (Cohen and Holder-Webb 2006). More recently, management accounting scholars have added to this foundation by examining social phenomena, such as norms (Altenburger 2017; Cardinaels and Yin 2015; Chen et al. 2017; Tayler and Bloomfield 2011) or contexts that impact perceptions of fairness (Cohen et al. 2007; Douthit and Stevens 2015; Guo et al. 2017; Zhang 2008). Other research has examined the impact of specific management or internal controls on honesty (Brink et al. 2017; Davidson and Stevens 2013; Hannan et al. 2010; Liu et al. 2015; Newman 2014; Rankin et al. 2008). These and other studies have found that management and internal controls, by themselves or in combination, impact reporting behavior in ways that might not be anticipated. Our measures can help explain some of these findings and allow us to dig deeper into how much of a change in behavior is due to contextual factors versus individual motivations.

With regard to individual differences, prior research notes distinctions between participants' preferences for honesty and/or preferences for appearing honest (e.g., Church et al. 2014; Evans et al. 2001; Hannan et al. 2006; Hao and Houser 2017). We contribute to this research stream by capturing these preferences *ex ante*. Some studies examine one or more of the dark triad traits of Machiavellianism, narcissism, and psychopathy (Ham et al. 2017; Majors 2016; Murphy 2012). Others, such as Blanthorne and Kaplan (2008), examine the effects of ethical beliefs on honest reporting. Our internal motivation measure is likely to capture such beliefs.

Hypothesis development for predictive validity

Self-determination theory helps us develop our hypotheses to test our measures' predictive validity. As described earlier, it theorizes that those who are higher in internal motivations will perform an act simply for its own sake. With regard to the act in question—reporting honestly—this implies that individuals higher in internal motivations to report honestly are expected to report more honestly than those who are lower in internal motivations. As a trait which is not closely tied to external stimuli, this tendency should hold across contexts. This leads to our first hypothesis that confirms the predictive validity of our internal measure:

HYPOTHESIS 1 (internal). *Individuals who are higher in internal motivations to report honestly will report more honestly compared to those who are lower in such motivations.*

Self-determination theory also suggests that those higher in external motivations will react to external stimuli. In other words, individuals who are higher in external motivations are not as likely

to report honestly without some reward for reporting honestly or consequence (e.g., punishment) for misreporting. Individuals who are lower in external motivations will not react as strongly to rewards or consequences because they are not motivated as much by such stimuli. Thus, in the absence of management controls, we expect no difference in reporting behavior between those who are higher or lower in external motivations. But when management controls designed to reduce misreporting are introduced, we expect individuals who are higher in external motivations to reduce their misreporting levels more than those who are lower in external motivations. This leads to our second set of hypotheses that confirms the predictive validity of our external measure:

HYPOTHESIS 2a (external). *In the absence of controls designed to reduce misreporting, individuals who are higher in external motivations to report honestly will report no more or less honestly than those who are lower in such motivations.*

HYPOTHESIS 2b (external). *In the presence (versus absence) of controls designed to reduce misreporting, individuals who are higher in external motivations to report honestly will reduce misreporting more than those who are lower in such motivations.*

Research questions with regard to the effect of controls

We now move to an important nuance with regard to management controls and their impact on individuals who are higher in internal motivations. First, the presence versus absence of controls should have a greater influence on those who are lower in internal motivations because individuals with higher motivations would have reported more honestly in their absence. However, as elaborated below, those higher in internal motivations may not react as expected depending on the nature of the controls. Thus, we formulate our general expectation concerning the effect of controls on internal motivations as a research question, as follows:

RESEARCH QUESTION 1 (internal). *Do individuals who are higher in internal motivations to report honestly differ in their propensity to report more honestly in response to the presence (versus absence) of controls from those who are lower in such motivations?*

When looking more closely at the impact of management controls on individuals who are higher in internal motivations, two theories suggest differing reactions, depending upon how the control is perceived. The first is cognitive evaluation theory, a subtheory of self-determination theory, and the second is motivation crowding theory.

Self-determination theory argues that individuals have three basic needs, for (i) competence, (ii) relatedness, and (iii) autonomy. To the extent individuals perceive the presence of these, they can function optimally. However, behavioral predictions change if these needs are not met. Cognitive evaluation theory focuses on the need for competence and autonomy, and experiments related to this theory investigate the effects of rewards and feedback on internal motivations. In essence, perceived threats to one's own competence or autonomy can affect one's natural internal motivation, potentially leading to actions *contrary* to what one would otherwise do (Deci et al. 1999; Ryan and Deci 2000).²⁰

The idea that intrinsic motivations may be undermined by external interventions has also been recognized within the economics literature using motivation crowding theory (Festre and Garrouste 2015; Frey 1994, 1997; Frey and Jegen 2001; Frey and Oberholzer-Gee 1997). According to motivation crowding theory, in the absence of intrinsic motivations, external interventions are generally assumed to simply have a disciplining effect (Frey 1994, 1997; Frey and Jegen 2001). However, in

20. Not all literature agrees with cognitive evaluation theory (e.g., see Kunz and Pfaff (2002) for a discussion of this theory as it relates to performance evaluations).

the presence of intrinsic motivations, the effect of external interventions will depend on whether those interventions are perceived to be supportive or controlling (Frey 1994, 1997; Frey and Jegen 2001). If the intervention is perceived to be supportive, the internal motivations are “crowded in,” which is theorized to enhance the external intervention. If it is perceived to be controlling, the internal motivations are “crowded out,” the results of which are theorized to depend upon the relative strength of the two effects. If the intrinsic motivation is sufficiently undermined by the external intervention, the disciplining effect will be entirely canceled out and the overall result will be negatively affected (Frey 1994, 1997; Frey and Jegen 2001).

Based upon these two theories, we argue that management controls designed to punish misreporting, rather than controls designed to reward honest reporting, could be perceived as controlling or threatening the autonomy of individuals who are higher in internal motivations. Put another way, controls that punish misreporting would crowd out internal motivations. By their nature, controls that punish misreporting are negative and are more likely to be seen by individuals who are already motivated to report honestly as an affront, because they impair self-determination. In contrast, management controls that acknowledge internal motivations by rewarding honest reporting could be perceived as supportive; such controls may foster individuals’ self-esteem. Given that the freedom to act is not affected, the sense of self-determination would be enhanced. Accordingly, controls that reward honest reporting would crowd in internal motivations. Overall, the specific effect of any control would depend on its nature and whether it is perceived as being controlling or supportive. While we can speculate about the possible side effects of controls, we cannot confirm *ex ante* how individuals will perceive them. Thus, we examine the more general research question, as follows:

RESEARCH QUESTION 2 (internal). *Do individuals who are higher in internal motivations to report honestly respond differently to controls that reward honest reporting versus controls that punish misreporting?*

Experiment 1

We administered our first experiment to Sample 2, discussed in section 2. We developed a generic task for MTurk workers, in which we provided both the opportunity and financial incentive to misreport, and we measured actual reporting behavior. Because the theory behind our measures is general in nature, it is appropriate to use MTurk participants without any particular experience from the general population and ask them to complete a basic task (Brandon et al. 2014). Participants were randomly assigned to one of three conditions, a baseline and two manipulations (*Reward* and *Punish*, explained later in this section) intended to test the measures, as well as controls over misreporting. Our initial sample size was 681. We removed two participants who did not complete any of the manipulation check questions, 10 participants who failed the comprehension check about the payment scheme (that they could report any amount in increments of 0.25 up to US\$5.00) and 69 participants who failed the manipulation check for their particular condition.²¹ In summary, 12 percent failed at least one of the comprehension or manipulation check questions, reducing the sample size from 681 to 600. The demographics of both the full Sample 2 and reduced Sample 2 are reported in Table 3.

The baseline proceeded as follows. After completing our survey for the honesty measures, the participants provided demographic information, performed a distractor task, and then completed a 20-question multiple choice trivia quiz, in which they earned US\$0.25 for each correct answer. Once finished, the computer automatically calculated and displayed their earned payment. Participants were then told that they could report any amount, in increments of US\$0.25, up to a maximum of US\$5.00, and that they would receive whatever amount they reported, which is

21. Thirty-seven participants failed the manipulation check for the *Reward* condition, and 32 failed for the *Punish* condition.

indeed what they were paid. Because participants could actually misreport, they were reminded that their responses were anonymous. Participants were each paid a flat fee of US\$2.00 for completing the honesty survey, plus an average of US\$2.58 for the experimental portion, based on their reports. After reporting their earnings, participants were asked several comprehension and manipulation check questions.

We administered two manipulations, intended to test reactions to controls, based on our measures, and to operationalize the distinction between rewarding honesty and punishing dishonesty from prior research (Chen 2012; Chen et al. 2017; Christ et al. 2012; Salterio and Webb 2006).²² In *Reward*, we devised a way to potentially reward honest reporting, whereas in *Punish*, we devised a way to potentially punish misreporting. Our first manipulation (*Reward*) stated the following: “**The researchers will be pleased with those who choose to report honestly.** *Further, the researchers reserve the right to grant a special qualification to all participants who report honestly, allowing them privileged access to their future studies*” (bolding and italics used to draw attention to our manipulation). Our second manipulation (*Punish*) stated the following: “**The researchers will be pleased with those who choose to report honestly.** *Further, the researchers reserve the right to block all participants who misreport from participating in any of their future studies*” (bolding and italics used to draw attention to our manipulation).²³

Results of Experiment 1

We examined responses to two statements about standards, with responses on a Likert scale from 1 = “strongly disagree” to 7 = “strongly agree.” The first read: “According to my own standards, I should have reported honestly in this task.” As expected, high internals responded significantly higher than low internals (mean response of 6.86 (SD = 0.68) and 5.85 (SD = 1.51), respectively, significant at $p < 0.01$ one-tailed), whereas high externals did not respond significantly different from low externals (6.40 (SD = 1.19) versus 6.32 (SD = 1.35)).²⁴ The second stated: “When reporting the earned payment from this task, I believe others would have reported honestly.” We believe that high externals were more likely to agree, in keeping with the idea that they want to act as they perceive their peers would. The mean response from high (low) externals is 4.69, SD = 1.87 (4.35; SD = 1.97), significant at $p = 0.03$, two-tailed. Thus, it appears that the participants in our reduced sample appropriately attended to our study.

Tables 6 and 7 present the results using the reduced sample size of 600, with Table 5 providing variable definitions.²⁵ Panel A of Table 6 reports the means of our dependent variable, *DIFF*, broken down by a median split of both the high and low internals (externals). *DIFF* represents the proportion of misreporting relative to the total amount possible (Evans et al. 2001; Mayhew and Murphy 2014), with a range from 0 (honest reporting) to 100 (misreporting to the fullest

-
22. Our experiments actually contained four conditions. After the baseline, we created a manipulation that addressed the theme of our external measure: the desire to please others. In this condition, labeled *Pleased*, we added the following statement: “**The researchers will be pleased with those who choose to report honestly.**” We intentionally bolded this sentence to draw attention to it. This condition is neither a strong manipulation nor a control per se, because participants are anonymous and can still misreport with no external consequences. Results indicate no significant effect of this condition; therefore, we do not discuss it in this manuscript, but we report results including this condition (which are substantively the same) in the online supplement.
 23. In *Reward* and *Punish*, we say that we *reserve the right* to either reward honest reporting or punish misreporting. Although we do not actually carry out either, no deception is involved.
 24. For the ease of exposition, we refer to those higher in internal motivations to report honestly as “high internals” and those higher in external motivations to report honestly as “high externals.” We use a median split for this purpose. For robustness, we also tried splitting our sample into quartiles and examining only the upper and lower 25 percent. Results show larger mean differences between cells and *t*-test results are similar or stronger.
 25. We report results from Experiment 1 (Sample 2) using the reduced sample, because it represents a cleaner test of our manipulations. When conducting the same tests using the full sample, we find that all of our inferences remain the same in sign and in significance, with one exception; the interaction of *Internal* and *Punish* becomes insignificant in panel B of Table 7.

TABLE 5
Variable definitions

Variable	Definition
<i>DIFF</i>	$[(\text{Reported Earned} - \text{Actual Earned}) / (5 - \text{Actual Earned})] \times 100$; with range from 0 to 100 (Experiment 1 only); higher number signifies a higher degree of misreporting
<i>LIKELY_YOU</i>	Likelihood the participant would allocate the \$2 million to one of the other two projects (misallocate); with range from 0 = “not at all likely” to 100 = “extremely likely” (Experiment 2 only); higher number signifies a greater likelihood of misallocation
<i>LikelyPeer</i>	Likelihood according to the participant that someone else in that position would allocate the \$2 million to one of the other two projects (misallocate); with range from 0 = “not at all likely” to 100 = “extremely likely” (Experiment 2 only); higher number signifies a greater likelihood of misallocation
<i>External</i>	External motivations to report honestly; factor capturing four items
<i>Internal</i>	Internal motivations to report honestly; factor capturing nine items
<i>Reward</i>	Experiment 1: A statement that the researchers would be pleased with anyone who reports honestly plus a statement that the researchers reserve the right to grant exclusive access to future studies to those who report honestly Experiment 2: A statement that the company’s founder would be pleased to know his employees were honest plus a statement that top management has promoted employees who have demonstrated honesty
<i>Punish</i>	Experiment 1: A statement that the researchers would be pleased with anyone who reports honestly plus a statement that the researchers reserve the right to block access to future studies to those who misreport Experiment 2: A statement that the company’s founder would be pleased to know his employees were honest plus a statement that top management has passed over employees who have been dishonest
<i>Controls</i>	Aggregate variable of <i>Reward</i> and <i>Punish</i>
<i>Earned</i>	The amount of payment actually earned by the participant based on their answers to the trivia questions (Experiment 1 only)
<i>Age</i>	Participant’s age (Experiment 1 only)
<i>Male</i>	One if male, zero otherwise (Experiment 1 only)
<i>WorkExp</i>	Years of work experience, with range from one = “none” to six = “over 20 years”; higher number means more work experience (Experiment 1 only)
<i>EmpStat</i>	One if employed full-time, zero otherwise (Experiment 1 only)
<i>CountryBorn</i>	One if born in the U.S. or Canada, zero otherwise (Experiment 1 only)
<i>DiffCase</i>	Answers to “This case was difficult to complete” with range from -4 = “strongly disagree” to 4 = “strongly agree” (Experiment 2 only)
<i>AcctgKnow</i>	Answers to “How much accounting knowledge do you have?” with a range from 0 = “virtually none” to 100 = “a great deal” (Experiment 2 only)

extent possible). To aid in interpreting this table, we label each cell with a letter. The table also reports *DIFF* for the *Reward* and *Punish* conditions combined (titled *Controls*) for purposes of testing the predictive validity of our measures in the presence of controls. The lowest level of misreporting is found in Cell C, the *Punish* condition for high internals while the highest level of misreporting is in Cell F, the *Baseline* for low internals.

TABLE 6
Descriptive statistics of first experiment, using reduced Sample 2

Panel A: Means of misreporting (DIFF), by median split of internal and external motivation		Baseline n = 224	Reward n = 185	Punish n = 191	Controls n = 376	Total n = 600
<i>DIFF by median split of Internal motivations</i>						
High Internal	A: n = 111	B: n = 94	C: n = 95	D: n = 189	E: n = 300	
	23.93	7.54	5.52	6.53	12.97	
	(42.13)	(25.33)	(22.09)	(23.72)	(32.83)	
Low Internal	F: n = 113	G: n = 91	H: n = 96	I: n = 187	J: n = 300	
	38.52	27.85	14.35	20.92	27.55	
	(47.47)	(44.43)	(33.37)	(39.62)	(43.52)	
<i>DIFF by median split of External motivations</i>						
High External	K: n = 96	L: n = 93	M: n = 89	N: n = 182	O: n = 278	
	30.26	18.82	7.15	13.11	19.03	
	(45.08)	(16.24)	(24.20)	(32.23)	(37.98)	
Low External	P: n = 128	Q: n = 92	R: n = 102	S: n = 194	T: n = 322	
	32.06	16.24	12.41	14.22	21.31	
	(45.79)	(37.18)	(31.86)	(34.45)	(40.24)	
Totals	31.29	17.53	9.96	13.69	20.26	
	(45.40)	(37.33)	(28.60)	(33.35)	(39.20)	
Panel B: Predictions and Welch's t-test results						
Prediction	Mean difference	Standard error	95% confidence interval	Welch's t	Significance	Cohen's D
Testing predictive validity						
H1: E < J	-14.58	3.14	-20.76 to -8.40	-4.63	0.01*	0.38
H2a: K = P	-1.81	6.13	-13.89 to 10.30	-0.29	0.77	0.04
H2b: (K-N) > (P-S)	-0.70	6.13	-12.78 to 11.38	-0.11	0.55*	0.02
Testing the effect of management controls on high internals						
RQ1: (A-D) vs. (F-I)	-0.19	5.99	-12.01 to 11.62	-0.03	0.97	0.01
RQ2: (A-B) vs. (A-C)	-2.02	5.66	-13.17 to 9.12	-0.36	0.72	0.05

(The table is continued on the next page.)

TABLE 6 (continued)

Panel C: Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11
(1) <i>DIFF</i>											
(2) <i>Internal</i>	-0.29										
(3) <i>External</i>	-0.01	0.06									
(4) <i>Reward</i>	-0.05	-0.01	0.04								
(5) <i>Punish</i>	-0.18	0.02	0.02	-0.46							
(6) <i>Earned</i>	-0.09	0.01	-0.01	-0.06	0.03						
(7) <i>Age</i>	-0.14	0.23	-0.20	-0.03	-0.02	0.21					
(8) <i>Male</i>	0.07	-0.22	0.07	-0.03	0.05	0.12	-0.21				
(9) <i>WorkExp</i>	-0.12	0.20	-0.14	0.05	-0.06	0.16	0.72	-0.10			
(10) <i>EmpStat</i>	0.05	-0.03	0.14	0.05	0.01	0.06	-0.12	0.19	0.19		
(11) <i>CountryBorn</i>	0.09	-0.07	-0.02	0.04	-0.02	0.02	-0.03	0.03	-0.01	0.02	

Notes: Panel A reports the mean (SD) of *DIFF* in each lettered cell. Internals are delineated by a median split at >8.27 ($n = 300$) for high, and ≤ 8.27 ($n = 300$) for low. Externals are delineated by a median split at >5.25 ($n = 322$) for high, and ≤ 5.25 ($n = 278$) for low. The sample mean for *Internal (External)* is 7.78 (5.16). We use continuous *Internal* and *External* variables in our regression analyses. The *Controls* column is a combination of *Reward* and *Punish*. Panel B: Bolded rows signify that the hypothesis or research question is supported. * indicates a one-tailed directional hypothesis test. All other tests are two-tailed. Panel C: Bolded numbers are significant at $p < 0.05$, two-tailed. Pearson (parametric) correlations are reported in the bottom triangle; Spearman (non-parametric) correlations are reported in the top triangle. See Table 5 for variable definitions.

We perform Welch's *t*-tests of our hypotheses by comparing cells. Our predictions and results, along with confidence intervals and Cohen's *D* (effect sizes) are reported in Table 6, panel B. Hypothesis 1 is supported in that high internals report more honestly than low internals ($p = 0.01$, one-tailed). We tested the same hypothesis regarding high versus low internals in the absence of controls only (untabulated) and found the difference to be significant ($p = 0.01$, one-tailed), also supporting Hypothesis 1. Hypothesis 2a is supported, as there is no significant difference in misreporting between high and low externals in the absence of controls ($p = 0.77$, two-tailed). Hypothesis 2b is not supported, as we find no evidence that high externals reduce their misreporting more than low externals in the presence (versus absence) of controls designed to reduce misreporting ($p = 0.55$, one-tailed). Exploring our first research question, we find there is no significant difference between high and low internals in terms of their propensity to report more honestly in response to the presence (versus absence) of controls ($p = 0.97$, two-tailed). The results pertaining to our second research question also indicate no evidence that high internals respond differently to controls that reward honesty versus those that punish misreporting ($p = 0.72$, two-tailed). Overall, the effect sizes for our significant results are in the range of small to medium, with the biggest effect size due to the presence of controls versus the baseline for internal motivations (Cohen 1988).²⁶

Before performing a regression analysis, we conduct a correlation analysis (see Table 6, panel C) to identify control variables. Consistent with prior literature (Mayhew and Murphy 2014; Murphy 2012), there is a negative correlation between *Earned* and *DIFF*, suggesting that those who earned more misreported less. We also find that age, gender, work experience, current work status, and country are significantly correlated. These correlations suggest that those who misreported more tended to be younger (consistent with the moral development literature (Kohlberg 1969; Rest 1986)), male, to have less work experience, to be employed full-time, and to be from the United States or Canada. We include these variables in our regressions, the results of which are reported in Table 7, with *DIFF* as the dependent variable.²⁷

Although the *t*-tests use a median split of our two motivation factors, our OLS regressions employ *Internal* and *External* as continuous variables. Table 7, panel A, reports the results of our regression, where the *Reward* and *Punish* condition are aggregated into *Controls*, while panel B reports results using a separate dummy variable for each control. As expected and consistent with the *t*-test, *Internal* is significant and negative ($p < 0.01$, one-tailed). This supports Hypothesis 1 and suggests that high internals misreport less than low internals. Also consistent with the *t*-test of Hypothesis 2a, *External* is not significant, suggesting that there are no significant differences in reporting behavior by those higher or lower in external motivations in the absence of controls ($p = 0.63$, two-tailed). Though this supports Hypothesis 2a, we caution against overinterpreting an insignificant result. Our findings with regard to Hypothesis 2b are also consistent with our *t*-tests, as they do not support the hypothesis. We similarly find an insignificant result from the interaction of *External* and *Controls* (panel A; $p = 0.26$, one-tailed) as well as *External* and each of *Reward* (panel B; $p = 0.18$, one-tailed) or *Punish* ($p = 0.39$, one-tailed).

Our research questions delve further into understanding how our internal measure interacts with controls. Regarding Research Question 1, we first establish with a *t*-test that high internals respond to the presence (versus absence) of controls by misreporting less (untabulated, $p = 0.02$, two-tailed). However, the lack of a significant interaction between *Internal* and *Controls* suggests that there is no

26. We follow Cohen (1988) and classify an effect size as small if $d = 0.20$, medium if $d = 0.50$, and large if $d = 0.80$. Basically, effect size is a more general comparison of means. If two means differ by 0.20 SD, then the difference is small, and if the means differ by 0.80 SD or more, then the difference is large.

27. Because some of our control variables are correlated with each other (e.g., *Age* and *WorkExp*), we tested our regression models for multicollinearity. Variance inflation factor scores for our independent and control variables were all well below the standard cutoff of 10 (with an average score of 1.45). Thus, we are confident that multicollinearity is not impacting our results.

TABLE 7

Experiment 1: Results of linear regression on *DIFF* (with robust standard errors)**Panel A:** Management controls aggregated

$$DIFF = \beta_0 + \beta_1 \text{Internal} + \beta_2 \text{External} + \beta_3 \text{Controls} + \beta_4 \text{Internal} \times \text{Controls} + \beta_5 \text{External} \times \text{Controls} \\ + \beta_6 \text{Earned} + \beta_7 \text{Age} + \beta_8 \text{Male} + \beta_9 \text{WorkExp} + \beta_{10} \text{EmpStat} + \beta_{11} \text{CountryBorn} + \varepsilon$$

Variable	Hypotheses: Predictions	Standardized coefficient	<i>t</i> -statistic	Significance
Intercept			4.40	<0.01
Internal	H1: –	–0.28	–3.44	<0.01*
<i>External</i>	H2a: ns	–0.04	–0.46	0.64
<i>Controls</i>		–0.36	–1.21	0.23
<i>Internal</i> × <i>Controls</i>	RQ1: ?	0.08	0.27	0.78
<i>External</i> × <i>Controls</i>	H2b: –	0.08	0.66	0.26*
Earned	–	–0.09	–2.32	0.01*
<i>Age</i>		–0.02	–0.37	0.71
<i>Male</i>		0.01	0.17	0.87
<i>WorkExp</i>		–0.06	–0.88	0.38
<i>EmpStat</i>		0.06	1.46	0.15
CountryBorn		0.07	5.55	<0.01

Panel B: Management controls disaggregated (test of RQ2)

$$DIFF = \beta_0 + \beta_1 \text{Internal} + \beta_2 \text{External} + \beta_3 \text{Reward} + \beta_4 \text{Punish} + \beta_5 \text{Internal} \times \text{Reward} \\ + \beta_6 \text{Internal} \times \text{Punish} + \beta_7 \text{External} \times \text{Reward} + \beta_8 \text{External} \times \text{Punish} + \beta_9 \text{Earned} \\ + \beta_{10} \text{Age} + \beta_{11} \text{Male} + \beta_{12} \text{WorkExp} + \beta_{13} \text{EmpStat} + \beta_{14} \text{CountryBorn} + \varepsilon$$

Variable	Hypotheses: Predictions	Standardized coefficient	<i>t</i> -statistic	Significance
Intercept			4.44	<0.01
Internal	H1: –	–0.28	–3.42	<0.01*
<i>External</i>	H2a: ns	–0.04	–0.49	0.63
<i>Reward</i>		–0.06	–0.17	0.86
Punish		–0.71	–2.30	0.02
<i>Internal</i> × <i>Reward</i>	RQ2: ?	–0.23	–0.82	0.41
<i>Internal</i> × <i>Punish</i>	RQ2: ?	0.44	1.59	0.11
<i>External</i> × <i>Reward</i>	H2b: –	0.12	0.90	0.18*
<i>External</i> × <i>Punish</i>	H2b: –	0.03	0.29	0.39*
Earned	–	–0.08	–2.13	0.02*
<i>Age</i>		–0.01	–0.21	0.83
<i>Male</i>		0.01	0.23	0.82
<i>WorkExp</i>		–0.07	–1.11	0.27
<i>EmpStat</i>		0.07	1.57	0.12
CountryBorn		0.07	5.04	<0.01

Notes: In panel A, $R^2 = 0.15$; $F = 10.09$ (significance < 0.01); number of observations: 600. In panel B, $R^2 = 0.17$; $F = 8.71$ (significance < 0.01); number of observations: 600. * indicates one-tailed. The dependent variable captures misreporting; a positive (negative) coefficient represents less (more) honest reporting. Significance for directional hypotheses are reported as one-tailed; the remainder are two-tailed. The bolded items are significant at $p \leq 0.05$. See Table 5 for variable definitions.

difference in how high versus low internals respond to the presence of controls (panel A; $p = 0.78$, two-tailed). With regard to the effects of different types of controls on reporting behavior, Research Question 2, panel B suggests a positive interaction between *Internal* and *Punish*, although it is only marginally significant if interpreted as a directional expectation ($p = 0.11$, two-tailed). This suggests that the disciplining effect of this type of control—one that punishes misreporting—has a diminished effect on high internals. The interaction of *Internal*×*Reward* is not statistically significant ($p = 0.41$, two-tailed). Thus, we find that high internals do not respond the same way to controls that reward honest reporting versus those that punish misreporting.²⁸

Overall, the results from this experiment establish predictive validity of our internal measure, as well as highlight an unintended consequence of a punishing control on the reporting behavior of high internals. Results also provide limited evidence of the predictive validity of our external measure.²⁹

Experiment 2

We operationalized similar constructs within a professional context by administering a second experiment through Prolific to individuals having management experience.³⁰ Recent research shows that Prolific respondents are more diverse but provide data that is of the same or greater quality than MTurk workers (Peer et al. 2017). Participants were each paid a flat fee of £4.00. With permission, we adapted the case from Chen et al. (2012) and Cohen et al. (2000). In our adaptation, participants were asked to think of themselves as plant managers within a large company, Pure Marine, which provides waste water treatment solutions to municipal and industrial markets. Specifically, they are one of several plant managers within the Membrane Group of Pure Marine, reporting to the Group VP. They were asked to indicate the likelihood that they, and the likelihood that someone else, would misallocate a \$2 million technical improvement expenditure. Allocating it to the correct project, K(3), would cause the participant to miss the target revenue and lose a substantial bonus. Alternatively, they could allocate the \$2 million to one of two newer multiyear projects that have enough slack to absorb this cost and still likely meet future targets. Relevant case facts indicated that the reason this choice was available is because of a poor cost accounting system. Even though the company policy clearly stated that costs should be allocated to the project on which they are incurred, the Group VP does not care if costs are misallocated and is more focused on current work being performed.³¹ The dependent variable, *LIKELY_YOU*, captures their likelihood of misreporting by asking the following: “How likely are you to allocate

-
28. We also performed a logistic regression using the data from Experiment 1 (Sample 2) with misreporting dichotomized as zero or one, to capture the likelihood of misreporting. Results are similar to those reported in Table 7. Hypothesis 1 and Hypothesis 2a are supported; high internals are 32 percent less likely to misreport compared to low internals ($p < 0.01$), while there is no significant difference in misreporting for high externals. In comparison to panel B, we find that the interaction between the *Reward* condition and our two measures has significant but opposite effects on the likelihood of misreporting. Compared to the *Baseline* condition, *Internals* were 21 percent less likely to misreport in the *Reward* condition, whereas *Externals* were 27 percent more likely to misreport in the *Reward* condition. These results respectively provide partial support for RQ2 but not for Hypothesis 2b. For completeness, we also conducted the logistic regression using the full sample and found similar results, except that the interaction between *External* and *Reward* as well as the interaction between *Internal* and *Reward* are insignificant.
29. For robustness we ran our analysis of Experiment 1 (Sample 2) after dropping MTurk participants who (i) did not have at least part-time employment and (ii) had not taken some university level courses. We find no significant differences related to our inferences of Table 6, panel B. The only noticeable difference is in the coefficient for *Internal*×*Punish*, which becomes insignificant in Table 7. All other variables remain unchanged.
30. Prolific, located in the United Kingdom, is similar to MTurk, except that when individuals sign up to be a Prolific participant, they complete a series of demographic questions, which allows researchers to recruit specific groups based on those demographics. We recruited individuals who had management experience.
31. We pilot tested this case with practicing managers who indicated the scenario was realistic.

the remaining \$2 million Technical Improvement expenditure for K(3) to either of the other projects that started this year?” Answers could range from 0 (“not at all likely”) to 100 (“extremely likely”).

The experiment proceeded as follows. After completing the honesty measures, participants were asked an extended series of background questions, the purpose of which was twofold: (i) to gather the usual demographics and (ii) to act as a distractor. For example, we asked them to describe the issues facing their organization and their role within their organization. Next, participants read the case and indicated both what they and what someone else would do in the situation. We asked the question in two ways to measure social desirability bias (Chung and Monroe 2003); however, we use the answer to what they, themselves, would do as our dependent variable. Our factors represent individual motivations, something emanating from inside oneself; thus, the relevant measure is their answer to what they would do, not what they think someone else would do. To capture the social desirability bias, we include participants’ reported likelihood that someone else in their position would misallocate (*LikelyPeer*) as a control variable in our regression. Finally, participants completed a post-experiment questionnaire almost identical to the one in the first experiment.

Similar to Experiment 1, we randomly assigned participants to one of three conditions. The baseline simply states that the presence of the company’s founder looms large and that employees are regularly reminded of his legacy of providing high-quality products. In the *Reward* condition, we add the following: “**One of the key values he espoused was honesty, and employees are continually reminded of how he would be pleased to know that his employees were honest. Company management has acted on this value in the past, most notably when it comes to promotions. Top management has promoted employees who have demonstrated honesty**” (bolding used to draw attention to our manipulation). This parallels *Reward* of our first experiment, which says that the researchers reserve the right to grant special access to those who report honestly. However, it differs because instead of offering an actual potential reward, it involves a hypothetical situation. Finally, *Punish* ends with “**Company management has acted on this value in the past, most notably when it comes to promotions. Top management has passed over employees who have been dishonest.**” Again, this parallels *Punish* in the first experiment, which indicates that the researchers reserve the right to block future participation for those who misreport. However, it is again hypothetical.

Though this experiment is analogous to the first experiment, it is distinct in two respects. First, it attempts to capture an actual business situation and uses participants with management experience. Second, the dependent variable, *LIKELY_YOU*, is similar to *DIFF* but is based on what participants would do in this situation versus making an actual reporting decision that is tied to their pay. Thus, the participants do not face the same incentives.

Results of Experiment 2

We administered the experiment to 224 participants who have management experience, using Prolific. The sample is reduced to 126 after deleting the following participants: 28 who failed one of two comprehension checks regarding the decision choices, 40 who failed the manipulation check question for their condition, 10 who indicated that they had no management experience (a requirement for this experiment), and 20 who took less than one minute to read the case facts. In summary, nearly 44 percent of participants failed at least one manipulation or comprehension check question, reducing the sample size from 224 to 126.

The average management experience of participants is eight years. The majority of participants are from the United Kingdom (54 percent), while most of the rest are from the United States (42 percent). Other demographic data shows 72 percent completed university, 82 percent have full-time employment, and 72 percent of participants have more than 11 years of work experience. Using a chi-square test to compare Sample 3 (Experiment 2) to Sample 2 (Experiment 1), we find that participants are more highly educated ($p < 0.01$) and more likely to be currently employed ($p < 0.01$), both of which are not surprising, given our intent to recruit individuals with management experience.

Similar to our first experiment, we asked about standards to ensure that high internals find their own standards to be important. Participants responded on a Likert scale from 1 = “strongly disagree” to 7 = “strongly agree” to the following statement: “According to my own personal standards, I should have allocated the entire \$2 million to project K(3)”. High internals’ mean response (5.57; SD = 2.15) is higher than low internals’ (5.33; SD = 1.62) but is not significantly different ($p = 0.24$, one-tailed). As expected, high externals did not significantly differ in their response from low externals (mean responses of 5.47 (SD = 1.77) and 5.43 (SD = 2.02), respectively).

Tables 8 and 9 present the results with the reduced sample.³² Table 8 reports the mean likelihood that participants, themselves, would misallocate the \$2 million, broken down by median splits of high versus low *Internal (External)* motivations, similar to the first experiment. As with Table 6, we label each cell to aide in interpreting our tests and results and combine the *Reward* and *Punish* conditions (*Controls*) for purposes of testing the predictive validity of our measures in the presence of controls. As panel A indicates, consistent with the first experiment, the prevalence of misallocation is highest (54 percent) among low internals in the baseline (Cell F). Unlike the first experiment, the lowest prevalence of misallocation, 27 percent, is by low externals in the *Punish* condition (Cell R).

Similar to our analysis in the first experiment, panel B reports the results of Welch’s *t*-tests, confidence internals, and Cohen’s *D*. As in our first experiment, Hypothesis 1 and Hypothesis 2a are supported. High internals report more honestly than low internals ($p = 0.06$, one-tailed), and they report more honestly than low internals in the absence of controls only (untabulated, $p = 0.02$ one-tailed). There is no significant difference in misreporting for high versus low externals in the baseline ($p = 0.37$, two-tailed), supporting Hypothesis 2a. We also find the same non-support for Hypothesis 2b as in the first experiment, as there is no evidence that high externals reduce their misreporting more than low externals in the presence (versus absence) of controls ($p = 0.18$, two-tailed). Concerning Research Question 1, the propensity to report more honestly in response to the presence of controls is less pronounced for high internals than for low internals ($p = 0.08$, two-tailed), unlike Experiment 1 in which there was no significant difference. This effect is attributable to high internals actually misreporting more (though not significantly so) in the presence versus absence of controls, while low internals misreported less. For Research Question 2, we find no difference in reporting behavior for high internals when we compare *Reward* versus *Punish* ($p = 0.13$, two-tailed). As reported by the Cohen’s *D*, our results are in the small to medium range (Cohen 1988).

Correlation analyses (see Table 8, panel C) report the expected correlations between *LIKELY_YOU* and *Internal* and *External*. We find that *DiffCase* (reported difficulty of the case), *AcctgKnow* (level of accounting knowledge), and *LikelyPeer* are positively and significantly correlated with *LIKELY_YOU*, so we include them in our regression analyses.³³

Table 9 reports the results of our linear regression analysis. *LIKELY_YOU* is the dependent variable, because it relates most directly to our factors (indicating what “you” would do). As with the first experiment, we run two different regressions using the continuous *Internal* and *External* variables. Panel A reports the results when *Reward* and *Punish* are combined, while panel B reports the results using a separate dummy variable for each control. *Internal* is again negative and significant ($p < 0.01$, one-tailed), supporting Hypothesis 1. Hypothesis 2a is again supported by the insignificance of the *External* variable ($p = 0.24$, two-tailed), though we

32. Performing the same analyses with the full Sample 3, we find that our inferences remain the same in sign and significance for panel A (panel B) of Table 9 with the following exceptions: *External* is positive and significant in both panels, the interaction between *Internal* and *Controls* is positive and significant in panel A, while *Reward* is negative and significant in panel B.

33. As in our analysis of Experiment 1 (Sample 2), we performed a test of multicollinearity. The average VIF score for our independent and control variables is 1.17, well below the standard threshold of 10.

TABLE 8
Descriptive statistics of second experiment, using reduced Sample 3

Panel A: Mean likelihood that the participant would misallocate (<i>LIKELY_YOU</i>), by median split of internal and external motivation						
	<i>Baseline</i> <i>n</i> = 41	<i>Reward</i> <i>n</i> = 38	<i>Punish</i> <i>n</i> = 47	<i>Controls</i> <i>n</i> = 85	Total <i>n</i> = 126	
<i>LIKELY_YOU</i> by median split of <i>Internal</i> motivations						
High <i>Internal</i>	A: <i>n</i> = 23 32.35 (33.97)	B: <i>n</i> = 16 49.31 (41.59)	C: <i>n</i> = 24 33.83 (33.60)	D: <i>n</i> = 40 40.03 (33.45)	E: <i>n</i> = 63 37.22 (33.57)	
Low <i>Internal</i>	F: <i>n</i> = 18 54.39 (33.17)	G: <i>n</i> = 22 41.59 (33.06)	H: <i>n</i> = 23 44.91 (34.53)	I: <i>n</i> = 45 43.29 (33.47)	J: <i>n</i> = 63 46.46 (33.50)	
<i>LIKELY_YOU</i> by median split of <i>External</i> motivations						
High <i>External</i>	K: <i>n</i> = 10 51.90 (40.08)	L: <i>n</i> = 21 53.86 (29.58)	M: <i>n</i> = 29 46.62 (35.18)	N: <i>n</i> = 50 49.66 (32.82)	O: <i>n</i> = 60 50.03 (33.77)	
Low <i>External</i>	P: <i>n</i> = 31 38.84 (33.29)	Q: <i>n</i> = 17 33.71 (33.12)	R: <i>n</i> = 18 27.39 (29.59)	S: <i>n</i> = 35 30.46 (31.05)	T: <i>n</i> = 66 34.39 (32.15)	
Totals	42.02 (35.00)	44.84 (32.41)	39.26 (34.15)	41.75 (33.30)	41.84 (33.72)	
Panel B: Predictions and Welch's <i>t</i> -test results						
Prediction	Mean difference	Standard error	95% confidence interval	Welch's <i>t</i>	Significance	Cohen's <i>D</i>
Testing predictive validity						
H1: E < J	-9.24	5.98	-21.07 to -2.59	-1.55	0.06*	0.28
H2a: K = P	13.06	14.01	-43.09 to 16.96	0.93	0.37	0.37
H2b: (K-N) > (P-S)	-6.14	14.01	-36.16 to 23.89	-0.44	0.67*	0.18

(The table is continued on the next page.)

TABLE 8 (continued)

Panel B: Predictions and Welch's <i>t</i> -test results								
Prediction	Mean difference	Standard error	95% confidence interval	Welch's <i>t</i>	Significance	Cohen's <i>D</i>		
Testing the effect of management controls on high internals								
RQ1: (A-D) vs. (F-I)	-18.78	10.55	-40.11 to 2.56	-1.78	0.08	0.56		
RQ2: (A-B) vs. (A-C)	-15.48	10.02	-35.65 to 4.69	-1.55	0.13	0.46		
Panel C: Correlation matrix								
	1	2	3	4	5	6	7	8
(1) <i>LIKELY_YOU</i>								
(2) <i>Internal</i>	-0.21							
(3) <i>External</i>	0.24	0.12						
(4) <i>Reward</i>	0.06	-0.08	0.13					
(5) <i>Punish</i>	-0.06	0.04	0.05	-0.51				
(6) <i>DiffCase</i>	0.22	0.12	0.20	0.11	0.01			
(7) <i>AcctgKnow</i>	0.21	0.05	0.15	0.12	-0.12	0.01		
(8) <i>LikelyPeer</i>	0.29	0.18	0.09	0.04	-0.11	-0.01	0.08	

Notes: Panel A reports the mean (SD) of *LIKELY_YOU* within each cell. Internals are delineated by a median split at >8.27 (*n* = 63) for high, and ≤8.27 (*n* = 63) for low. External are delineated by a median split at >5.5 (*n* = 60) for high, and ≤5.5 (*n* = 66) for low. The sample mean for *Internal (External)* is 7.97 (5.63). We use continuous *Internal* and *External* variables in our regression analyses. The *Controls* column is a combination of *Reward* and *Punish*. Panel B: Bolded rows signify that the hypothesis or research question is supported. * indicates a one-tailed directional hypothesis test. All other tests are two-tailed. Panel C: Bolded numbers are significant at *p* < 0.05, two-tailed. Pearson (parametric) correlations are reported in the bottom triangle; Spearman (nonparametric) correlations are reported in the top triangle. See Table 5 for variable definitions.

TABLE 9

Experiment 2: Results of linear regression on *LIKELY_YOU* (with robust standard errors)**Panel A: Management controls aggregated**

$$LIKELY_YOU = \beta_0 + \beta_1 \text{Internal} + \beta_2 \text{External} + \beta_3 \text{Controls} + \beta_4 \text{Internal} \times \text{Controls} \\ + \beta_5 \text{External} \times \text{Controls} + \beta_6 \text{DiffCase} + \beta_7 \text{AcctgKnow} + \beta_8 \text{LikelyPeer} + \varepsilon$$

Variable	Hypotheses: Predictions	Standardized coefficient	<i>t</i> -statistic	Significance
Intercept			2.54	0.01
Internal	Hypothesis 1: –	–0.46	–4.11	<0.01*
<i>External</i>	Hypothesis 2a: ns	0.20	1.18	0.24
<i>Controls</i>		–0.62	–1.07	0.29
<i>Internal</i> × <i>Controls</i>	Research Question 1: ?	0.60	1.15	0.25
<i>External</i> × <i>Controls</i>	Hypothesis 2b: –	–0.03	–0.09	0.47*
DiffCase		0.24	2.95	<0.01*
AcctgKnow		0.17	2.37	0.02
LikelyPeer		0.32	4.41	<0.01

Panel B: Management controls disaggregated (test of RQ2)

$$LIKELY_YOU = \beta_0 + \beta_1 \text{Internal} + \beta_2 \text{External} + \beta_3 \text{Reward} + \beta_4 \text{Punish} \\ + \beta_5 \text{Internal} \times \text{Reward} + \beta_6 \text{Internal} \times \text{Punish} + \beta_7 \text{External} \times \text{Reward} \\ + \beta_8 \text{External} \times \text{Punish} + \beta_9 \text{DiffCase} + \beta_{10} \text{AcctgKnow} + \beta_{11} \text{LikelyPeer} + \varepsilon$$

Variable	Hypotheses: Predictions	Standardized coefficient	<i>t</i> -statistic	Significance
Intercept			2.49	0.01
Internal	Hypothesis 1: –	–0.46	–4.06	<0.01*
<i>External</i>	Hypothesis 2a: ns	0.20	1.15	0.25
<i>Reward</i>		–1.00	–1.53	0.13
<i>Punish</i>		–0.15	–0.25	0.81
<i>Internal</i> × <i>Reward</i>	Research Question 2: ?	1.00	1.77	0.08
<i>Internal</i> × <i>Punish</i>		0.12	0.23	0.82
<i>External</i> × <i>Reward</i>	Hypothesis 2b: –	–0.08	–0.23	0.41*
<i>External</i> × <i>Punish</i>		–0.01	–0.02	0.49*
DiffCase		0.25	3.03	<0.01*
AcctgKnow		0.18	2.30	0.02
LikelyPeer		0.32	4.20	<0.01

Notes: In panel A, $R^2 = 0.31$; $F = 10.29$ (significance < 0.01); number of observations: 126. In panel B, $R^2 = 0.32$; $F = 7.94$ (significance < 0.01); number of observations: 126. * indicates one-tailed. The dependent variable captures misreporting; a positive (negative) coefficient represents less (more) honest reporting. Significance for directional hypotheses are reported as one-tailed; the remainder are two-tailed. The bolded items are significant at $p \leq 0.05$. See Table 5 for the variable definitions.

again caution against overinterpreting this result. Our test of the effects of controls on high externals (Hypothesis 2b) is directionally consistent with our predictions but not significant ($p = 0.47$, one-tailed).

Similar to our first experiment, we find no significant results in the regression for Research Question 1 (*Internal*×*Controls* is insignificant, $p = 0.25$, two-tailed). A *t*-test also shows that high internals' misreporting is not statistically related to the presence (versus absence) of controls (untabulated, $p = 0.38$, two-tailed). However, we do find different results regarding Research Question 2. Whereas results from experiment one report a marginally significant and positive

interaction between *Internal* and *Punish*, our Experiment 2 results, reported in panel B of Table 9, show a marginally significant and positive interaction of *Internal* and *Reward* ($p = 0.08$, two-tailed). That is, the disciplining effect of a control in which management promotes those who report honestly is diminished for high internals. The interaction between *Internal* and *Punish* is not significant ($p = 0.82$, two-tailed).

To better understand the difference in results for Research Question 2 between the simple *t*-tests and our regression, we return to the theory informing our research question. Both cognitive evaluation theory and motivation crowding theory support the notion that a control perceived to be supportive (controlling) should positively (negatively) interact with internal motivations. In post-experiment questions, we asked participants if they felt constrained, consistent with the theory that high internals will feel more constrained under controls perceived to be a threat. Results of *t*-tests show a marginally significant increase in feelings of constraint by high internals in the *Reward* condition versus the baseline ($p = 0.10$, one-tailed), whereas no significant difference is found between the *Punish* condition and the baseline. Thus, even though this differs from our results in Experiment 1, the post-experiment question provides partial support that the perception of the control led to the reporting behavior, consistent with theory. This explanation emphasizes the importance of how employees (or subjects) perceive controls, not how managers (or researchers) perceive them.

Other findings not hypothesized include a significant positive result of *DiffCase* ($p < 0.01$, two-tailed) and *AcctgKnow* ($p = 0.02$, two-tailed). Regarding *DiffCase*, we asked those who agreed with this statement ($n = 51$) why they thought the case was difficult. Approximately 73 percent expressed difficulty in putting themselves into the hypothetical situation or feeling conflicted or guilty for having to make such a decision. We consider these reasons reasonable, as it is the nature of the case. Specifically, participants were confronted with conflicting objectives and an ethical choice, so it is normal for them to indicate that it was difficult. However, about 22 percent stated that the case was too complex or they had trouble comprehending it. The significant and positive *AcctgKnow* finding is troubling, because it suggests that the more accounting knowledge participants had, the more they misreported. Nevertheless, it corresponds with the view that ethics education is lacking in the accounting field (Armstrong 1987; Cohen and Holder-Webb 2006; Cohen and Pant 1989; Liu et al. 2012).³⁴ Finally, we find that *LikelyPeer* is positive and significant, but the coefficient is less than one, suggesting that respondents may be exhibiting a social desirability bias in their responses.³⁵

Summary results of experiments

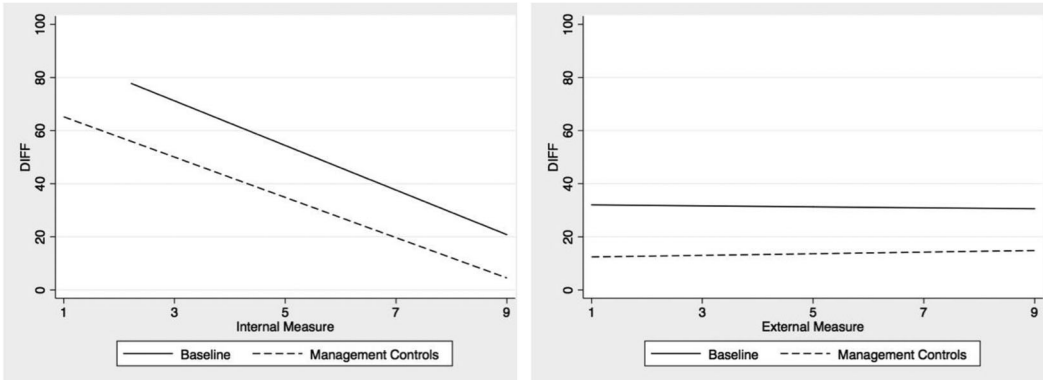
To summarize our predictive validity analyses, both experiments support our hypothesis that high internals report more honestly than low internals (Hypothesis 1). The two experiments also confirm that high externals report no differently than low externals in the absence of controls (Hypothesis 2a). However, we do not find support in either experiment for Hypothesis 2b that

34. We performed additional analyses for Experiment 2 (Sample 3) by comparing results for those in the lower and upper third of *AcctgKnow*. For the lower third, *Internal*, *External*, and *Reward* are negative and significant. The interactions between *Internal* and *Reward* as well as *Internal* and *Reward* are both positive and significant. However, for the upper third, only *External* is significant, but it is positive, suggesting that those with high levels of accounting knowledge and that are high externals will misreport more. The fact that *Internal* is not significant for those with higher accounting knowledge may be because those with lower accounting knowledge rely more heavily on internal motivations, while those with more accounting knowledge have become desensitized to the slippery slope of misreporting, and thus do not rely on their internal motivations, but instead on their knowledge of accounting.

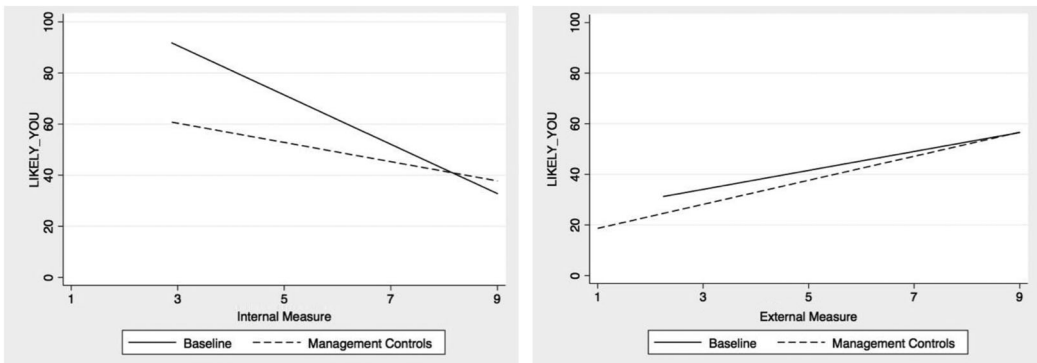
35. Similar to Chen et al. (2012), we performed additional analysis. We measured participants' perceived ethicality by taking the difference between *LIKELY_YOU* and *LikelyPeer*. Results suggest that the higher an individual scores on our internal measure, the more unethical (i.e., larger difference) they find the misreporting choice. However, this view of ethicality is diminished when internals are exposed to the *Reward* condition. Results related to the *External* measure are insignificant.

Figure 1 These graphs plot the dependent variable in each experiment as a function of the level of internal and external motivations. Panel A: Degrees of misreporting (*DIFF*) in Experiment 1. Panel B: Likelihood of misallocation (*LIKELY_YOU*) in Experiment 2.

A Degrees of misreporting (*DIFF*) in the Experiment 1



B Likelihood of misallocation (*LIKELY_YOU*) in the Experiment 2



high externals will report more honestly in the presence versus absence of controls than low externals. Finally, results pertaining to our two research questions about how high internals react to controls are mixed. For Research Question 1, we find no significant difference in the propensity to report more honestly in response to the presence (versus absence) of controls between high and low internals in the first experiment, whereas in Experiment 2, we find a marginally significant difference in the *t*-tests but not in the regression analysis. Exploring how high internals respond to controls that are framed as rewarding versus punishing (Research Question 2) also shows mixed results between the experiments. In the first, we find diminished reactions of high internals in the *Punish* condition, whereas in the second, we found the same effect in the *Reward* condition. When *Reward* and *Punish* are aggregated, these effects are offset. We thus contend that both experiments support the predictive validity of our internal measure and provide limited support of the predictive validity of our external measure.

We believe there is a potential explanation for the differing reactions by high internals to the *Reward* versus *Punish* conditions in our two experiments. It is likely that Experiment 1 was taken more seriously by MTurk workers than Experiment 2 by Prolific participants. We tested actual reporting behavior with the former and hypothetical judgments with the latter. Our rewards and punishments may have been perceived as directly impacting MTurk workers' future income. According to Berg (2016), 37 percent of American MTurk workers indicate that crowdwork is

TABLE 10
Summary of research suggestions and practice implications

Panel A: Regarding the measure of *internal* motivations to report honestly

Suggestions for research	Implications for practice
<ul style="list-style-type: none"> • Administer this measure when examining any (mis)reporting behavior, to control for individual differences. • Examine related accounting phenomena using this measure, such as corporate disclosure decisions, propensity to whistle-blow, or aggressive reporting. • Examine interactions between this measure and other contextual factors, such as perceptions of fairness. • Examine how this measure interacts with moral foundations (Graham et al. 2013). How do individuals higher in internal motivations respond when a moral foundation is threatened? • Study this measure with participants over time to verify that it is trait-like (stable) versus state-like (changeable). • Examine optimal contracting solutions for employees who are higher in this measure. • Examine the association between this measure and moral identity (Aquino and Reed 2002). 	<ul style="list-style-type: none"> • Administer this measure to employees who report performance (e.g., division heads, management). • Survey employees higher in internal motivations regarding their view of controls. Do they view them as threatening, controlling or constraining? • Use more stringent management controls and monitoring for workforces lower in internal motivations. • More supportive management controls and monitoring might be more cost effective for workforces higher in internal motivations.

Panel B: Regarding the measure of *external* motivations to report honestly

Suggestions for research	Implications for practice
<ul style="list-style-type: none"> • Develop and test additional measures of external motivations based on other situational factors (e.g., need for respect or reputation). • Compare other measures of external motivations to this one and establish predictive validity. • Explore the relation between the dark triad traits and external motivations. • Examine interactions between this measure and other contextual factors such as perceptions of fairness (Cohen et al. 2007), and issue-contingent factors (Jones 1991). • Cross high/low internal and external motivation measures and examine impact on reporting decisions. • Conduct a field experiment using this and other measures of external motivation to better understand how employees who are higher/lower in external motivations react to different controls. 	<ul style="list-style-type: none"> • None suggested until predictive validity is established.

their primary source of income and 94 percent have had their work rejected or not paid. These statistics support the notion that our control manipulations, especially the *Punish* condition, were likely quite strong for this group of participants, unlike in the second experiment where participants' decisions had no bearing on their payment. It is also possible that manager participants who are high in internal motivations perceived our *Reward* and *Punish* conditions differently than anticipated. Perhaps they reacted to an inconsistency in the second experiment in which the case

facts allude to the participant's boss being unconcerned about misallocations; yet, corporate management promotes based on honest allocations. Post-experiment questions provide some evidence that the perceptions of high internals to the *Reward* control was a factor in their decisions, consistent with the theories behind these hypotheses. An important takeaway from these results is that employees who are higher in internal motivations can and do react differently to controls seen as too controlling or threatening. This unintended consequence should be a warning for management and those charged with governance.

Figure 1 plots the dependent variables of both experiments as a function of increasing internal or external motivations to report honestly. For Experiment 1, panel A reports *DIFF*, first for internal motivations, then external motivations. For Experiment 2, panel B reports *LIKELY_YOU*, first for internal motivations, then external motivations. As expected, panel A (Experiment 1) clearly shows that misreporting declines as internal motivations increase, whereas misreporting levels do not change based on the level of external motivations to report honestly. The effect of controls is that they reduce misreporting overall but not as a function of either internal or external motivations. Panel B (Experiment 2) also shows that, as internal motivations increase, misreporting decreases. However, it also reveals that misreporting actually increases as external motivations increase. The effect of controls is also different in the second experiment. Although controls appear to have an overall effect but not an interaction with motivation, they interact with internal motivations at the highest end of the motivation continuum. This suggests that controls perceived to be constraining may impact those highest in internal motivations to report honestly. This is useful information for managers who want to survey their workforce to determine their motivations and then design controls that fit them.

4. Conclusion

The overarching objective of this paper is to create and validate two measures that capture internal and external motivations to report honestly, as individual traits (see Table 4 for the individual items in the measure and part A of the online supplement for the entire measure). To accomplish this, we collected data from more than 2,300 participants using three sources—students, MTurk, and Prolific—and we conducted two complementary experiments to test predictive validity. Our experiments examined real reporting decisions, involving members from the general public, and hypothetical accounting decisions, involving individuals with management experience.

Our measure of internal motivations passes our validity tests. It is internally consistent, with Cronbach's alphas ranging from 0.91 to 0.95 for all our samples (see part G of the online supplement for details). It is robust to all our convergent and divergent validity tests, and it predicts behavior in both experiments. We also provide evidence that individuals who are higher in internal motivations to report honestly can react negatively to controls perceived to be threatening, constraining, or controlling. This should be of concern to management and those charged with governance as they consider the types of management controls to implement within their organizations. We suggest that managers measure their employees' levels of internal motivations to report honestly, as well as their views of the current management controls. For workforces with mostly high internals, management may choose to implement controls that are perceived to be supportive rather than threatening.

While our external measure is internally consistent, with Cronbach's alphas ranging from 0.77 to 0.93 for all our samples (see part G of the online supplement for details), and while it passes our convergent and divergent validity tests, results from both experiments provide only limited evidence of its predictive validity. Thus, we do not recommend using this measure in practice until it has been validated more thoroughly, but we recommend further research. One possible avenue is to explore different themes beyond the wish to please others, such as the need for respect or reputation (Rom and Conway 2018; Vonasch et al. 2018). We also suggest comparing the character traits in the dark triad with external motivations to report honestly in the presence of different external stimuli (O'Boyle et al. 2012).

Overall, our results directly add to a research stream that examines honesty in reporting. For example, Evans et al. (2001) and Church et al. (2014) consider preferences for honesty, while Hannan et al. (2006) discuss findings relating to managers' wishes to appear honest. By actually measuring individuals' desire to *be* honest (internal motivations) and their desire to *appear* honest (external motivations), we extend this literature. Our findings suggest that there is a way to measure an individual's utility for honest reporting as an individual motivation, which can be utilized in future studies. Our evidence also suggests that different management controls will garner different reactions depending upon one's level of internal motivations.

Our measures provide insights to other research streams as well. For instance, greenwashing, when a company purports to be environmentally friendly even though it is not (Berrone et al. 2017; Lyon and Maxwell 2011), may be due to top management being higher in external motivations. Findings that associate changes in corporate disclosures with changes in top managements' incentives (e.g., Hui and Matsunaga 2015 in bonuses and Levy et al. 2018 in changes in personal litigation risk) may be related to external motivations to report honestly. The notion of cheap talk in some accounting literature (Bowlin et al. 2015; Qu 2013) may be more prevalent among high externals than high internals.

We suggest that future research employ these measures to better understand their interactions with other contextual factors. For example, how do our measures interact with the concept of fairness? On the one hand, preferences for fairness (Luft 1997) can be considered to be part of one's utility function. If so, how closely does this value relate to our internal motivation to report honestly? On the other hand, fairness is also a matter relating to the perception of outcomes (Cohen et al. 2007). Therefore, how does perceived fairness in certain situations interact with our measures to predict reporting behavior? More generally, how do the moral foundations of care, fairness, loyalty, respect for authority, and purity (Graham et al. 2013) interact with our measures in situations in which honest reporting or misreporting threaten any of those foundations? Overall, these general research ideas can be applied to multiple streams within the accounting literature besides management accounting. Any research that considers honesty in reporting, such as corporate disclosures, auditing or tax, can gain insights using our measures.

We also call for more research incorporating our measures in order to better understand their interactions with management controls. We summarize some ideas and suggestions in Table 10. Other research questions include (i) Do our measures apply equally to aggressive reporting as they do to misreporting?; (ii) What can we learn from crossing high and low internals and externals? For example, are high internal / low externals likely to be the most honest reporters and high external / low internals the least honest reporters?; (iii) Are these measures stable over time, as expected for most character traits?; (iv) How closely related is the external motivation factor to the dark triad?; and (v) What are the optimal contracts for individuals high in internal (external) motivations?³⁶ There are many more potential research questions, and we believe this paper begins a fruitful line of research intended to better understand individuals' wish to *appear* honest versus *being* honest, and how these individual traits play out within organizations.

References

- Akerlof, G. A. 1982. Labor contracts as partial gift exchange. *Quarterly Journal of Economics* 97 (4): 543–69.
- Altenburger, M. 2017. The effect of injunctive social norms and dissent on budget reporting honesty. *Journal of International Accounting Research* 16 (2): 9–31.

36. The budgeting literature has identified and examined reporting behavior under different types of contracts including the trust contract and modified trust (hurdle) contract (Akerlof 1982; Church et al. 2012; Evans et al. 2001). These contracts could be reexamined in conjunction with our measures. Although we argue that a trust contract is likely optimal for high internals, this remains an empirical question.

- Aquino, K., and I. I. Reed. 2002. The self-importance of moral identity. *Journal of Personality and Social Psychology* 83 (6): 1423–40.
- Ariely, D. 2012. *The honest truth about dishonesty*. New York: HarperCollins.
- Armstrong, M. B. 1987. Moral development and accounting education. *Journal of Accounting Education* 5 (1): 27–43.
- Association of Certified Fraud Examiners (ACFE). 2018. *Report to the nations: 2018 Global study on occupational fraud and abuse*. Austin, TX: ACFE.
- Baiman, S., and B. Lewis. 1989. An experiment testing the behavioral equivalence of strategically equivalent employment contracts. *Journal of Accounting Research* 27 (1): 1–20.
- Barbuto, J. E., and R. W. Scholl. 1998. Motivation sources inventory: Development and validation of new scales to measure an integrative taxonomy of motivation. *Psychological Reports* 82 (3): 1011–22.
- Basham, B. 2011. Beware corporate psychopaths—They are still occupying positions of power. *Independent*, Thursday, December 20.
- Berg, J. 2016. *Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers*. Geneva, Switzerland: International Labour Organization (ILO).
- Berrone, P., A. Fosfuri, and L. Gelabert. 2017. Does greenwashing pay off? Understanding the relationship between environmental actions and environmental legitimacy. *Journal of Business Ethics* 144 (2): 363–79.
- Blanthorne, C., and S. Kaplan. 2008. An egocentric model of the relations among the opportunity to underreport, social norms, ethical beliefs, and underreporting behavior. *Accounting, Organizations and Society* 33 (7–8): 684–703.
- Bowlin, K. O., J. L. Hobson, and M. D. Piercey. 2015. The effects of auditor rotation, professional skepticism, and interactions with managers on audit quality. *The Accounting Review* 90 (4): 1363–93.
- Brandon, D. M., J. H. Long, T. M. Loraas, J. Mueller-Phillips, and B. Vansant. 2014. Online instrument delivery and participant recruitment services: Emerging opportunities for behavioral accounting research. *Behavioral Research in Accounting* 26 (1): 1–23.
- Brink, A. G., J. C. Coats, and F. W. Rankin. 2017. Deceptive superiors and budgetary reporting: An experimental investigation. *Journal of Management Accounting Research* 39 (3): 79–91.
- Cardinaels, E., and H. Yin. 2015. Think twice before going for incentives: Social norms and the principal's decision on compensation contracts. *Journal of Accounting Research* 53 (5): 985–1015.
- Chen, C. X., J. Nichol, and F. H. Zhou. 2017. The effect of incentive framing and descriptive norms on internal whistleblowing. *Contemporary Accounting Research* 34 (4): 1757–78.
- Chen, M. L. 2012. The effect of leader reward and punishment behaviors on subordinates' budget reports. *Engineering Economist* 57 (1): 41–54.
- Chen, Q., K. Kelly, and S. E. Salterio. 2012. Do changes in audit actions and attitudes consistent with increased auditor scepticism deter aggressive earnings management? An experimental investigation. *Accounting, Organizations and Society* 37 (2): 95–115.
- Christ, M. H., S. A. Emmett, S. L. Summers, and D. A. Wood. 2012. The effects of preventive and detective controls on employee performance and motivation. *Contemporary Accounting Research* 29 (2): 432–52.
- Chung, J., and G. S. Monroe. 2003. Exploring social desirability bias. *Journal of Business Ethics* 44 (4): 291–302.
- Church, B. K., R. L. Hannan, and X. Kuang. 2012. Shared interest and honesty in budget reporting. *Accounting, Organizations and Society* 37 (3): 156–67.
- Church, B. K., R. L. Hannan, and X. Kuang. 2014. Information acquisition and opportunistic behavior in managerial reporting. *Contemporary Accounting Research* 31 (2): 398–419.
- Churchill, G. A., Jr. 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16 (1): 64–73.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. New York: Routledge Academic.
- Cohen, J. R., and L. L. Holder-Webb. 2006. Rethinking the influence of agency theory in the accounting academy. *Issues in Accounting Education* 21 (1): 17–30.
- Cohen, J., G. B. Manzon, and V. L. Zamora. 2015. Contextual and individual dimensions of taxpayer decision making. *Journal of Business Ethics* 126 (4): 631–47.

- Cohen, J. R., and L. W. Pant. 1989. Accounting educators' perceptions of ethics in the curriculum. *Issues in Accounting Education* 4 (1): 70–81.
- Cohen, J. R., L. W. Pant, and D. J. Sharp. 2000. Project earnings manipulation: An ethics case based on agency theory. *Issues in Accounting Education* 15 (1): 89–103.
- Cohen, J. R., L. W. Pant, D. J. Sharp, and L. Holder-Webb. 2007. The effect of perceived fairness on opportunistic behavior. *Contemporary Accounting Research* 22 (4): 1119–38.
- Davidson, B. I., and D. E. Stevens. 2013. Can a code of ethics improve manager behavior and investor confidence? An experimental study. *The Accounting Review* 88 (1): 51–74.
- Deci, E. L., R. Koestner, and R. M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6): 627–68.
- Deci, E. L., A. H. Olafsen, and R. M. Ryan. 2017. Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior* 4 (1): 19–23.
- Deci, E. L., and R. M. Ryan. 2012. Self-determination theory. In *Handbook of theories of social psychology*, edited by P. A. M. Van Lange, A. W. Kruglanski, and E. T. Higgins, 416–37. Los Angeles: Sage.
- DeVellis, R. F. 1991. *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications.
- Devine, P. G., P. S. Forscher, A. J. Austin, and W. T. L. Cox. 2012. Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology* 48 (6): 1267–78.
- Dorminey, J. W., A. C. Fleming, and M. Kranacher. 2011. Beyond the fraud triangle. *Fraud Magazine*, September/October.
- Douthit, J. D., and D. E. Stevens. 2015. The robustness of honesty effects on budget proposals when the superior has rejection authority. *The Accounting Review* 90 (2): 467–93.
- Drake, A. R., L. J. Matuszewski, and F. Miller. 2014. The effect of personality traits and fairness on honesty in managerial reporting. *Advances in Management Accounting* 22: 43–69.
- Evans, J. H. III, R. L. Hannan, R. Krishnan, and D. V. Moser. 2001. Honesty in managerial reporting. *The Accounting Review* 76 (4): 537–59.
- Fehr, E., and K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114 (3): 817–68.
- Festre, A., and P. Garrouste. 2015. Theory and evidence in psychology and economics about motivation crowding out: A possible convergence? *Journal of Economic Surveys* 29 (2): 339–56.
- Fiolleau, K., T. Libby, and L. Thorne. 2018. Dysfunctional behavior in organizations: Insights from the management control literature. *Auditing: A Journal of Practice & Theory* 37 (4): 117–41.
- Frey, B. S. 1994. How intrinsic motivation is crowded out and in. *Rationality and Society* 6 (3): 334–52.
- Frey, B. S. 1997. *Not just for the money*. Cheltenham, UK: Edward Elgar Publications.
- Frey, B. S., and R. Jegen. 2001. Motivation crowding theory. *Journal of Economic Surveys* 15 (5): 589–611.
- Frey, B. S., and F. Oberholzer-Gee. 1997. The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review* 87 (4): 746–55.
- Gibbons, R. 1998. Incentives in organizations. *Journal of Economic Perspectives* 12 (4): 115–32.
- Gibson, R., C. Tanner, and A. F. Wagner. 2013. Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review* 103 (1): 532–48.
- Gneezy, U., B. Rockenbach, and M. Serra-Garcia. 2013. Measuring lying aversion. *Journal of Economic Behavior and Organization* 93: 293–300.
- Graham, J., J. Haidt, S. Koleva, M. Motyl, R. Iyer, S. P. Wojcik, and P. H. Ditto. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology* 47: 55–130.
- Graham, J., R. Iyer, B. A. Nosek, J. Haidt, S. Koleva, and P. H. Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology* 101 (2): 366–85.
- Grover, L. S. 2005. The truth, the whole truth and nothing but the truth: The causes and management of workplace lying. *Academy of Management Executive* 19 (2): 148–57.
- Guo, L., T. Libby, and K. Liu. 2017. The effects of vertical pay dispersion: Experimental evidence in a budget setting. *Contemporary Accounting Research* 34 (1): 555–76.

- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108 (4): 814–34.
- Haidt, J., J. Graham, and C. Joseph. 2009. Above and below left-right: Ideological narratives and moral foundations. *Psychological Inquiry* 20 (2): 110–9.
- Ham, C., J. Lang, N. Seybert, and S. Wang. 2017. CFO narcissism and financial reporting quality. *Journal of Accounting Research* 55 (5): 1089–135.
- Hannan, R., F. Rankin, and K. Towry. 2006. The effect of information systems on honesty in managerial reporting: A behavioral perspective. *Contemporary Accounting Research* 23 (4): 885–918.
- Hannan, R. L., F. W. Rankin, and K. L. Towry. 2010. Flattening the organization: The effect of organizational reporting structure on budgeting effectiveness. *Review of Accounting Studies* 15 (3): 503–36.
- Hao, L., and D. Houser. 2017. Perceptions, intentions, and cheating. *Journal of Economic Behavior & Organization* 133: 52–73.
- Hogan, C., Z. Rezaee, R. Riley, and U. Velury. 2008. Financial statement fraud: Insights from the academic literature. *Auditing: A Journal of Practice & Theory* 27 (2): 231–52.
- Hu, L., and P. M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1–55.
- Hui, K. W., and S. R. Matsunaga. 2015. Are CEOs and CFOs rewarded for disclosure quality? *The Accounting Review* 90 (3): 1013–47.
- Hurttt, R. K. 2010. Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory* 29 (1): 149–71.
- Iacobucci, D. 2010. Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology* 20 (1): 90–8.
- International Auditing and Assurance Standards Board (IAASB). 2009. The auditor's responsibilities relating to fraud in an audit of financial statements. International Standard on Auditing 240. New York: IAASB.
- Jones, J., D. Massey, and L. Thorne. 2003. Auditors' ethical reasoning: Insights from past research and implications for the future. *Journal of Accounting Literature* 22: 45–103.
- Jones, T. M. 1991. Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review* 16 (2): 366–95.
- Judd, S., E. O'Rourke, and A. Grant. 2018. Employee surveys are still one of the best ways to measure engagement. *Harvard Business Review*, March 14.
- Kohlberg, L. 1969. Stage and sequence: The cognitive-developmental approach to socialization. In *Handbook of Socialization Theory and Research*, edited by D. Goslin 347–480. Chicago: Rand McNally.
- Kunz, A. H., and D. Pfaff. 2002. Agency theory, performance evaluation, and the hypothetical construct of intrinsic motivation. *Accounting, Organizations and Society* 27 (3): 275–95.
- Leary, M. R. 1983. A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin* 9 (3): 371–5.
- Lee, K., and M. C. Ashton. 2004. Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research* 39 (2): 329–58.
- Lee, K., M. C. Ashton, J. Wiltshire, J. S. Bourdage, B. A. Visser, and A. Gallulcci. 2013. Sex, power, and money: Prediction from the dark triad and honesty-humility. *European Journal of Personality* 27 (2): 169–84.
- Levy, H., R. Shalev, and E. Zur. 2018. The effect of CFO personal litigation risk on firms' disclosure and accounting choices. *Contemporary Accounting Research* 35 (1): 434–63.
- Liu, C., L. J. Yao, and N. Hu. 2012. Improving ethics education in accounting: Lessons from medicine and law. *Issues in Accounting Education* 27 (3): 671–90.
- Liu, X., A. Wright, and J. Wu. 2015. Managers' unethical fraudulent financial reporting: The effect of control strength and control framing. *Journal of Business Ethics* 129 (2): 295–310.
- López-Pérez, R., and E. Spiegelman. 2013. Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics* 16 (3): 233–47.

- Lowe, E. A., and R. W. Shaw. 1968. An analysis of managerial biasing: Evidence from the company's budgeting process. *Journal of Management Studies* 5 (3): 304–15.
- Luft, J. L. 1997. Fairness, ethics and the effect of management accounting on transaction costs. *Journal of Management Accounting Research* 9: 199–216.
- Lyon, T. P., and J. W. Maxwell. 2011. Greenwash: Corporate environmental disclosure under threat of audit. *Journal of Economics & Management Strategy* 20 (1): 3–41.
- Maas, V. S., and M. Van Rinsum. 2013. How control system design influences performance misreporting. *Journal of Accounting Research* 51 (5): 1159–86.
- Majors, T. 2016. The interaction of communicating measurement uncertainty and the dark triad on managers' reporting decisions. *The Accounting Review* 91 (3): 973–92.
- Mayhew, B. W., and P. R. Murphy. 2014. The impact of authority on reporting behavior, rationalization and affect. *Contemporary Accounting Research* 31 (2): 420–43.
- Miles, A., and L. Upenieks. 2018. An expanded model of the moral self: Beyond care and justice. *Social Science Research* 72: 1–19.
- Murphy, K. 1993. *Honesty in the workplace*. Pacific Grove, CA: Brooks.
- Murphy, P. R. 2012. Attitude, Machiavellianism and the rationalization of misreporting. *Accounting, Organizations and Society* 37 (4): 242–59.
- Newman, A. 2014. An investigation of how the informal communication of firm preferences influences managerial honesty. *Accounting, Organizations and Society* 39 (3): 195–207.
- Nunnally, J. C., and R. H. Bernstein. 1994. *Psychometric theory*. New York: McGraw-Hill.
- O'Boyle, E. H., D. R. Forsyth, G. C. Banks, and M. A. McDaniel. 2012. A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology* 97 (3): 557–79.
- Paulus, D., and K. Williams. 2002. The dark triad of personality. *Journal of Research in Personality* 36 (6): 556–63.
- PCAOB. 2005. *Consideration of fraud in a financial statement audit*. AU Section 316. www.pcaobus.org.
- Peer, E., L. Brandimarte, S. Samat, and A. Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70: 153–63.
- Plant, E. A., and P. G. Devine. 1998. Internal and external motivation to respond without sexism. *Journal of Personality and Social Psychology* 75 (3): 811–32.
- Qu, H. 2013. How do market prices and cheap talk affect coordination? *Journal of Accounting Research* 51 (5): 1221–60.
- Rankin, F. W., S. T. Schwartz, and R. A. Young. 2008. The effect of honesty and superior authority on budget proposals. *The Accounting Review* 83 (4): 1083–99.
- Rest, J. R. 1986. *Moral development: Advances in research and theory*. New York: Praeger.
- Rom, S. C., and P. Conway. 2018. The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology* 74: 24–37.
- Rosenberg, M. 1965. *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Ryan, R. M., and E. L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55 (1): 68–78.
- Salterio, S. E., and A. Webb. 2006. Honesty in accounting and control: A discussion of “the effect of information systems on honesty in managerial reporting: A behavioral perspective.” *Contemporary Accounting Research* 23 (4): 919–32.
- Sandino, T. 2011. Developing good measures to advance management accounting and control research: A discussion of “corporate frugality: Theory, measurement, and practice.” *Contemporary Accounting Research* 28 (4): 1388–96.
- Snyder, M., and S. Gangestad. 1986. On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology* 51 (1): 125–39.
- Tangney, J. P., R. F. Baumeister, and A. L. Boone. 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality* 72 (2): 271–324.
- Taylor, W. B., and R. J. Bloomfield. 2011. Norms, conformity, and controls. *Journal of Accounting Research* 49 (3): 753–90.

- Trompeter, G., T. Carpenter, N. Desai, K. Jones, and D. Riley. 2013. A synthesis of fraud related research. *Auditing: A Journal of Practice & Theory* 32 (1): 287–321.
- Vonasch, A. J., T. Reynolds, B. M. Winegard, and R. F. Baumeister. 2018. Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science* 9 (5): 604–13.
- Zhang, Y. 2008. The effects of perceived fairness and communication on honesty and collusion in a multi-agent setting. *The Accounting Review* 83 (4): 1125–46.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1. Supporting Information