Moral Mercenaries:
The Paradoxical Effect of Pay-for-Performance on Unethical Tasks

by

Julian House

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Joseph L. Rotman School of Management
University of Toronto

© Copyright by Julian House 2015
Moral Mercenaries: The Paradoxical Effect of Pay-for-Performance on Unethical Tasks

Julian House
Doctor of Philosophy
Joseph L. Rotman School of Management
University of Toronto
2015

Abstract

While some organizations, like charities and government agencies, engage in prosocial behavior that upholds the highest prescriptive moral values in our society; others, like those in the tobacco, fossil fuel, gambling, and weapons industries, operate within the confines of the legal system while nevertheless behaving antisocially and violating core moral proscriptions against harming others. Unfortunately our understanding of the psychology that enables collective moral violations is still very limited because scholarship in behavioural business ethics has tended to focus on individual-level antecedents of unethical behaviour as opposed to the interpersonal and organizational factors that lead ordinarily ethical people astray. This dissertation attempts to rectify this imbalance by asking how it is possible to motivate employees to comply with instructions that violate society’s ethical principles. Given that money is generally considered the most powerful motivational tool available to most organizations, this research examines how different financial incentive structures motivate performance of dirty work that is technically legal but violates society’s ethical norms. Surprisingly, and contrary to its well-documented positive effect on the performance of ethically-neutral work tasks, I predict that pay-for-performance incentive structures will have a negative effect on performance of unethical tasks. Existing motivational theories are unable to make similar predictions, so I contribute to our understanding of unethical behavior on behalf of organizations by considering the embedded nature of ethical decision-making and argue that
pay-for-performance alters employees’ perceived employment relationships and thereby inhibits moral disengagement from societal norms.

Through a series of pilot studies I identify two appropriate experimental paradigms to test my hypotheses. The first reported experiment involves a novel paradigm in which online workers are recruited as part of a marketing study to compose social-media advertising messages in support of a tobacco company. The second experiment adapts an established bug-killing paradigm in which laboratory participants are asked to place either bugs or inanimate objects into a modified coffee grinder that they subsequently activate. In both experiments, participants are either paid for their performance of this moderately unethical task or receive a comparable flat fee. This dissertation reports the results of these experiments and discusses their scholarly contribution.
Acknowledgments

I would like to express gratitude to Chen-Bo Zhong and Sanford DeVoe for their support as supervisors of my dissertation. I am also grateful to the other members of my supervisory committee, Stéphane Côté and Nina Mazar. The advice, input, and perspective of these scholars was critical to this process. I would also like to acknowledge the contributions of The Michael Lee-Chin Family Institute for Corporate Citizenship and the Social Sciences and Humanities Research Council of Canada to this research.
# Table of Contents

Introduction ................................................................................................................................. 1  
Motivating Destructive Compliance ............................................................................................ 6  
Pay-for-performance Inhibits Authority Relationships and Moral Disengagement ................. 10  
Methodology Overview .............................................................................................................. 16  
Pilot Test 1 ............................................................................................................................... 17  
  Participants and Procedure .................................................................................................... 18  
  Manipulation and Measures ................................................................................................. 20  
  Results ................................................................................................................................. 25  
Pilot Test 2 ............................................................................................................................... 26  
  Participants and Procedure ................................................................................................ 27  
  Manipulation and Measures ............................................................................................... 29  
  Results ................................................................................................................................. 33  
  Discussion ............................................................................................................................ 37  
Pilot Test 3 ............................................................................................................................... 38  
  Participants and Procedure ................................................................................................ 39  
  Manipulation and Measures ............................................................................................... 41  
  Results ................................................................................................................................. 44  
Experiment 1 ............................................................................................................................ 46  
  Participants and Procedure ................................................................................................ 46  
  Manipulation and Measures ............................................................................................... 48  
  Results ................................................................................................................................. 55  
  Discussion ............................................................................................................................ 64  
Experiment 2 ............................................................................................................................ 72  
  Participants and Procedure ................................................................................................ 73  
  Manipulation and Measures ............................................................................................... 74  
  Results ................................................................................................................................. 78  
  Discussion ............................................................................................................................ 80  
General Discussion .................................................................................................................. 82  
References ................................................................................................................................ 87
List of Tables

Table 1. Correlation matrix of variables from Pretest 2. .......................................................... 106
Table 2. Correlation matrix of variables from Pilot Test 3. ......................................................... 110
Table 3. Correlation matrix of variables from Experiment 1. ..................................................... 112
Table 4. Correlation matrix of variables from Experiment 2. ..................................................... 117

List of Figures

Figure 1. Proposed theoretical model. .......................................................................................... 104
Figure 2. Graph of the extent to which participants lied in Pilot Test 1. ................................. 105
Figure 3. Objective performance on advertising task in Pilot Test 2 by pay structure and moralization of tobacco advertising. ................................................................. 107
Figure 4. Objective performance on advertising task in Pilot Test 2 by pay structure and internal moral identity centrality. ......................................................... 107
Figure 5. Perception of employment relationship as authority ranked in Pilot Test 2 by pay structure and moralization of tobacco advertising. ................................. 108
Figure 6. Laboratory set-up in Pilot Test 3 and Experiment 2. ................................................. 109
Figure 7. Performance on grinding task in Pilot Test 3 by pay structure, ethicality of task, and gender of research assistant. ................................................................. 111
Figure 8. Sense of personal responsibility for the advertising task in Experiment 1 by pay structure and moralization of tobacco advertising. ................................................................. 113
Figure 9. Number of advertising messages written by participants in Experiment 1 by pay structure and internal moral identity. ................................................................. 114
Figure 10. Participants’ quality-weighted performance scores on the advertising task in Experiment 1 by pay structure and internal moral identity. ................................................................. 115
Figure 11. Participants’ sense of guilt after completing the advertising task in Experiment 1 by pay structure and internal moral identity. ................................................................. 116
Figure 12. Performance on grinding task in Experiment 2 by pay structure, ethicality of task, and participants’ internal moral identity. ................................................................. 118

List of Appendices

Appendix A. Laboratory protocol for Pilot Test 3 and Experiment 2. ......................................... 119
Appendix B. Coding protocol for Experiment 1. ........................................................................ 131
Introduction

Although many organizations operate business models that are arguably morally neutral, providing goods and services for customers within the market, it is not uncommon for organizations to engage in (un)ethical enterprise. While some organizations, like charities and government agencies, engage in prosocial behavior that upholds the highest prescriptive moral values in our society; others, like those in the tobacco, fossil fuel, gambling, and weapons industries, operate within the confines of the legal system while nevertheless behaving antisocially and violating core moral proscriptions against harming others. Unfortunately our understanding of the psychology that enables collective moral violations is still very limited because scholarship in behavioural business ethics (see De Cremer & Tenbrunsel, 2012 for an excellent review) has tended to focus on individual-level antecedents of unethical behaviour as opposed to the interpersonal and organizational factors that lead ordinarily ethical people astray (Moore & Gino, 2013). Moreover, much of the extant research into interpersonal influences on moral decision making has focussed on the effects of power (see Lammers, Galinsky, Dubois, & Rucker, 2015, for a recent review of this research). Therefore while we may have a crystalizing understanding of the moral behaviour of powerful leaders, we know relatively less about what factors encourage their relatively powerless subordinates to adhere to their (un)ethical instructions and why (Treviño, den Nieuwenboer, & Kish-Gephart, 2014).

Rectifying this imbalance is important because when it comes to unethical behavior we know that social influence can encourage otherwise normal individuals to commit shockingly inhumane acts (Haney, Banks, & Zimbardo, 1973; Milgram, 1963). Moreover, the consequences of collective action are generally larger than the sum of individuals’ capabilities. For example, tobacco companies’ products kill about 30 times as many Americans than does homicide every
year (Murphy, Xu, & Kochanek, 2013; U.S. Surgeon General, 2014). Of course such scale is only possible when large numbers of individuals diligently coordinate their efforts, but very little is known about unethical behavior carried out on behalf of organizations (Umphress & Bingham, 2011). How it is possible to motivate employees to comply with instructions that violate society’s ethical principles? Given that money is generally considered the most powerful motivational tool available to most organizations (Locke, Feren, McCaleb, Shaw, & Denny, 1980), this research examines how different financial incentive structures motivate performance of dirty work that is technically legal but violates society’s ethical norms (Ashforth & Kreiner, 1999; Ingram & Silverman, 2013). Surprisingly, and contrary to its well-documented positive effect on the performance of ethically-neutral work tasks, I predict that pay-for-performance incentive structures will have a negative effect on performance of unethical tasks. Existing motivational theories are unable to make similar predictions, so I contribute to our understanding of unethical behavior on behalf of organizations by considering the embedded nature of ethical decision-making (Rai & Fiske, 2011) and argue that pay-for-performance alters employees’ perceived employment relationships and thereby inhibits moral disengagement from societal norms.

From the perspective of the two dominant theories within the employee incentive literature, motivating performance of any task, ethical or otherwise, is best achieved by making incentives contingent upon that performance. On the one hand, agency theory (Jensen & Meckling, 1976) builds on classical economic assumptions to view employees as agents whose preferences are distinct from those of their employers, and posits that the purpose of incentive systems is to more closely align agent and principal’s preferences. On the other, expectancy theory (Vroom, 1964) examines intrapsychic motivation as the product of beliefs about the
ability to perform, the value of any reward received for performance, and the instrumentality of performance for gaining that reward. According to both theories, by tightening the coupling between performance and valued rewards, performance-based financial incentives increase employees’ motivation to comply with employers’ instructions. Indeed, multiple meta-analyses have found pay-for-performance incentive structures to be one of the most powerful human resource interventions for increasing employee performance (Guzzo, Jette, & Katzell, 1985; Jenkins, Mitra, Gupta, & Shaw, 1998; Locke et al., 1980; Stajkovic & Luthans, 1998), and as a result they are frequently recommended by management scholars (e.g., Lawler, 2000) and have been widely adopted by many companies (U.S. Bureau of National Affairs, 1988). Thus, because agency and expectancy theories are agnostic about the ethicality of the task being incentivized, both would predict that performance-based incentives are the best way to motivate compliance with any request, including unethical ones.

Subsequent research, however, has shown that financial incentives may backfire for certain tasks; yet the theories which explain this reversal for prosocial tasks do not apply to the performance of unethical tasks. According to both self-determination theory (Deci & Ryan, 1985, 2000), and what economists refer to as motivation crowding (Frey, 1993, 2007), financial incentives can interfere with intrinsic motivation and the perception of normative pressures to perform a task. Thus, when it comes to tasks that employees find intrinsically interesting or pleasant, or to tasks which are normatively prescribed, such as behaving prosocially (Heyman & Ariely, 2004; Hossain & Li, 2014), pay-for-performance is likely to interfere with these pre-existing motivations and may reduce performance. Indeed, pay-for-performance has been found to have negative or null relationships with performance of volunteers (Frey & Goette, 1999; Gneezy & Rustichini, 2000), blood donors (Niza, Tung, & Marteau, 2013), and physicians
In the absence of these motivations, however, pay-for-performance should augment task performance by increasing the extrinsic motivation to obtain rewards. Therefore, because similar motivations do not typically drive unethical behavior, self-determination and motivation crowding theories should also predict that performance-based incentives will serve to increase unethical performance. After all, unethical behavior is counter-normative by definition, so it is unlikely to be motivated by conformity to normative pressures. Instead, it tends to be motivated by selfish desires for extrinsic rewards (Loewenstein, 1996), rather than intrinsic pleasure because healthy individuals empathically experience the suffering of victims (Blair, 1995; Cushman, Gray, Gaffey, & Mendes, 2012) and treat moral self-image as a critical component of their self-worth (Blasi, 1984; Dunning, 2007). Consequently, with few, if any, intrinsic or normative motivations with which to interfere, all extant theoretical perspectives appear poised to predict that pay-for-performance incentive structures are an effective means of encouraging compliance with unethical instructions.

Paradoxically, in this paper I will argue that the current understanding of incentives cannot accurately predict how incentive structures will effect unethical behavior without considering the process of moral disengagement and the embedded nature of ethical decision-making. Bandura (1986, 1999) described moral disengagement as a set of cognitive strategies that people use to disregard their personal moral standards and avoid self-condemnation. These self-serving interpretations of situations and behavior allow individuals to benefit from the proceeds of unethical conduct without negatively updating their moral self-image, and may also be useful as arguments for protecting or enhancing one’s moral reputation among others (Haidt, 2001, 2007). As a result, when situational features facilitate construal of behavior as morally acceptable (Mazar, Amir, & Ariely, 2008), or the perceived diffusion of responsibility amongst
other co-actors (Diener, Fraser, Beaman, & Kelem, 1976; Luhan, Kocher, & Sutter, 2009), people are more likely to display self-interested unethical behavior. In fact, at the insistence of an authority figure, people will commit surprising acts of inhumanity even without direct personal benefit because of an “agentic shift” of responsibility for their behavior to the authority (Milgram, 1974). More recent research further suggests that obedience to such an authority may be seen as an ethical obligation in its own right (Graham et al., 2011), which may also encourage construal of such behavior as ethical.

What this research seeks to contribute is evidence that performance-based incentives can disrupt the process of moral disengagement by changing the way that employees perceive their employment relationship. This line of argument is related to the observation that employees’ ethical decisions can be shaped by circumstances which differentially activate alternate occupational identities (Leavitt, Reynolds, Barnes, Schilpzand, & Hannah, 2012), but is broader because it is not restricted to employees with disparate occupational identities (e.g., military medics who are both soldiers and healthcare professionals). I argue that the employer-employee relationship is itself inherently ambiguous: it can be viewed as either a hierarchical relationship between a superior and their subordinate, or as an exchange of remuneration for services between individuals in a market. Importantly, these different relationship construals imply different ethical obligations because moral psychology is fundamentally embedded within relationships (Rai & Fiske, 2011). While perceiving one’s employer as an authority who should be obeyed and who is ultimately responsible for decisions seems likely to promote moral disengagement, the perception of an exchange between individuals implies greater discretion and individual responsibility and thus may serve to inhibit this process. Thus performance-based incentives should make employees feel more independent and individually responsible, and
therefore less likely to morally disengage to comply with unethical instructions from their employer. I will test this hypothesis using two experiments in which participants are provided remuneration to promote tobacco products and kill (insects) at the behest of an authority figure.

**Motivating Destructive Compliance**

The focus of this research is on employee behavior that is consistent with organizational norms yet violates broader societal or hypernorms. Generally associated with implicit institutions necessary for human survival, hypernorms are ethical beliefs and values that enjoy broad global support (Donaldson & Dunfee, 1994, 1999), such as moral proscriptions against violence and stealing as well as the prescription to help those in need. Hypernorms can be differentiated from sets of local norms that are specific to a particular group or organization in order to identify four categories of deviant/conforming employee behavior (Warren, 2003). Destructive deviance is behavior that falls outside of both organizational and global norms, such as embezzlement and many counterproductive work behaviors (Gruys & Sackett, 2003). Constructive deviance describes behavior that violates organizational norms while upholding global norms, an example of this being whistle-blowing. Here I examine how financial incentives effect destructive conformity, or the adherence to organizational norms that violate hypernorms, such as working for a company that sells products or services which harm people. As explained above, this focus is theoretically different from previous work on pay-for-performance of constructive conformity, or behaviour which upholds both organizational and hypernorms, because there are no intrinsic and social motivations with which financial incentives should interfere. My focus on destructive conformity is also distinct from the small amount of theoretical (Vardi & Weitz, 2004; Vardi & Wiener, 1996) and empirical (Umphress & Bingham, 2010) work on unethical pro-organizational behaviour. Although these constructs
are related, because both involve unethical behaviour that is intended to benefit the organization, destructive compliance is more narrowly focused on organizationally normative behaviour, whereas unethical pro-organizational behaviour encompasses unethical behaviours “whether they are consistent with or defy organizational expectations” (Umphress & Bingham, 2011, p. 624). This narrower definition is theoretically significant because conformity to organizational norms and expectations is what makes organizing possible. Organizations only exist insofar as “there is a probability that certain persons will act in such a way as to carry out the order governing the organization” (Weber, 2013, p. 53). Motivating destructive compliance is thus necessary for the very existence of organizations with unethical business models and an important phenomenon to understand.

From the perspective of agency theory, the most effective means of motivating conformity to organizational norms and expectations, or compliance with specific instructions, is by making incentives contingent upon that conformity. This perspective draws upon the utilitarian philosophical tradition (Mill, 1879) and social exchange theory (Homans, 1958) in its assumptions that conformity and compliance, and the larger organizational structures they make possible, are the result of individuals’ rational decisions to maximize rewards and minimize punishments. Pay-for-performance thus increases compliance because compliant behavior facilitates employees’ utility maximization. Notably this perspective is consistent with the traditional explanation of unethical behavior: individuals succumb to unethical temptations when they calculate that the benefit of such behavior outweighs the expected cost of discovery and punishment (Becker, 1968; Kagan & Scholz, 1983). By enhancing the benefit of unethical behavior, performance-based pay should increase employees’ compliance with the unethical requests of their superiors according to agency theory.
The rational, cost-benefit explanation of unethical behavior has received only weak empirical support, however. Decades of research in criminology, for instance, has revealed that punishment severity and likelihood are among the weakest predictors of crime rates (Pratt & Cullen, 2005). Moreover, calculating the expected cost of unethical behavior seems unrealistically complicated for the limited cognitive capacity of the decision-maker when one expands punishment to include less easily quantified social sanctions. In place of this objective calculation, psychologists have suggested that individuals’ behavior is often guided by moral self-regard (Monin & Jordan, 2009), which serves as a heuristic sociometer estimate of others’ evaluation of our own moral status (Haidt, 2012). Consequently, because people place great value on their own moral self-regard (Blasi, 1984; Dunning, 2007), they often act in ways that are inconsistent with the cost-benefit model. For example, Mazar and colleagues (2008) have convincingly shown that people fail to maximize their personal benefit through cheating behavior, even when there is no risk of this behavior being discovered. Although impunity did increase cheating behavior in their experiments, it did so only slightly, presumably because the threat to participants’ moral self-image constrained their unethical behavior.

Of course, in addition to valuing their moral self-regard, people also value external rewards, and they often use flexible cognitive strategies to obtain the latter without sacrificing the former. According to social cognitive theory (Bandura, 1991), moral self-regulation occurs though a process of self-monitoring. When people judge that their behavior upholds their personal moral standards they maintain a positive self-regard, and when they judge their behavior as deviant they update their moral self-regard negatively. This process is far from objective, however, and because rewards are pleasant while lowering one’s moral self-regard is unpleasant, motivated reasoning often permits deviation from moral standards without self-
sanctioning. This process of avoiding self-regulation is called moral disengagement (Bandura, 1999) and it includes several different cognitive strategies, which are variously aimed at benignly reframing the action in question, disavowing personal responsibility, minimizing the negative consequences of the action, and devaluing the status of any victims. At the level of individual differences, the propensity to morally disengage is related to unethical behaviour at work (Moore, Detert, Treviño, Baker, & Mayer, 2012). However the likelihood that any particular individual will moral disengage, and which disengagement strategy they are likely to use, can also be influenced by situational constraints. For example, Mazar and colleagues (2008) also found that people are more likely to cheat when a medium of tokens separates their behavior from its proceeds, compared to participants whose cheating translated directly into cash. The researchers’ explanation for this behavior is that the additional transaction allowed for more room in which to frame one’s behavior as harmless: cheating for tokens seems less unsavory than cheating for cash. Likewise, Bandura, Underwood, and Fromson (1975) were able to influence the extent to which participants morally disengaged and harmed others by telling them that their influence over the magnitude of electric shocks delivered to a third party was either individually or jointly determined. When the pain caused to another person was jointly determined, moral responsibility was no longer individually held and this diffusion of responsibility facilitated more inhumane behaviour. In a final example, Wiltermuth (2011) has shown people are more likely to cheat when the proceeds are split with a third-party beneficiary because the consequence of benefiting another person allows the perpetrator to see their act as less immoral.

Returning to employee compliance with unethical instructions, the literature on moral self-regard and disengagement suggests that the normative justifications available to employees
within an organizational context may be important. Weber (2013) originally introduced the idea that organizational members are influenced to comply with the instructions of superiors according to “principles of domination.” No matter the kind of work an employee is instructed to perform, such principles provide them with a substantive rationale for complying, behavioral guidance, justifications, and interpretive frameworks for their actions as part of the organization (Biggart & Hamilton, 1984; Hamilton & Biggart, 1985). However, in the context of unethical tasks, these normative justifications may be especially important in order to facilitate moral disengagement and to protect employees’ moral self-regard so that they can more fully comply with instructions. Indeed, previous research into organizational corruption and dirty work has discussed how managers and socialization processes attempt to provide employees with ideological justifications for their unethical work (Ashforth & Anand, 2003; Ashforth, Kreiner, Clark, & Fugate, 2007). In the absence of such justifications, the stigma associated with dirty work is thought to interfere with employee commitment and performance (Ashforth & Kreiner, 1999), in much the same way that moral self-regard interferes with unethical profit maximization (Mazar et al., 2008). The normalizing strategies discussed in the dirty work literature also resemble some of the moral disengagement strategies proposed by Bandura (1999). For example, both describe strategies of reframing actions, recalibrating standards of comparison, and the disavowal of responsibility. Thus although financial incentives may be a predominant motivator in many jobs, when it comes to compliance with unethical requests, normative justifications which facilitate or inhibit moral disengagement may be relatively more important determinants of employee performance.

Pay-for-performance Inhibits Authority Relationships and Moral Disengagement
Why might pay-for-performance incentive structures interfere with the moral disengagement that makes compliance with unethical instructions more likely? I argue that such interference may occur because pay-for-performance changes the way in which employees view their relationships with their employers. Employees’ construal of their employment relationship is important, especially in the domain of unethical tasks, because just as economic decisions are embedded within relationships (Granovetter, 1985) so are ethical decisions (Fiske, 2002; Rai & Fiske, 2011). Consistent with the functional, sociometer view of moral self-regard discussed above, Rai and Fiske have argued that moral psychology more generally motivates relationship regulation. Moral emotions, such as compassion for in-group members, and reverence or fear of authority figures, motivate behaviors that help to sustain those relationships. Importantly, there are four basic relational models which people use to make sense of their relationships (Fiske & Haslam, 2005; Fiske, 1991), and much of the variance in moral behavior between and within individuals and cultures can be explained by the relationships which actors perceive to be relevant to particular decisions. For example, if someone perceives themselves to be in a communal-sharing relationship, such as a family or other close-nit in-group, they will be motivated to maintain unity and see provision of aid based on need, rather than merit, as a moral imperative. In contrast, if market pricing is perceived to govern a relationship, such as that between a buyer and seller, actors will be motivated to achieve outcomes proportional to inputs and see this equitable distribution as fair.

The relational model which appears poised to facilitate moral disengagement in favor of compliance with unethical instructions is authority ranking. When people see a relationship as characterized by a legitimate authority and his or her subordinates, their motivation is to preserve that hierarchy. Subordinates view respect, obedience and deference to the authority as
an ethical obligation. In turn, authorities feel a sense of pastoral duty such that in addition to leading, they also protect and assume responsibility for subordinates. Together, this sense of moral obligation to obey authority figures (Graham et al., 2011; Rai & Fiske, 2011) and the transfer of responsibility up the chain of command (Schultz, Jaggi, & Schleifer, 1987) constitute two cognitive strategies that make it easier for subordinates to morally disengage and simply follow orders. In the face of an instruction from an authority figure, subordinates may reframe a potentially unethical act as ethical by focusing on their sense of obligation to obey. Moreover, they may displace their sense of personal responsibility for the action onto the authority figure who issued the instruction. Thus as Milgram (1963, 1974) famously showed, and as has been replicated in contemporary American society (Burger, 2009), a large majority of ordinary people are willing to obey an authority figure who instructs them inflict pain on a stranger with potentially dangerous levels of electric shocks.

Because relationships do not possess objective features that determine to which relational model they are most similar, relationship perceptions may exert influence over moral psychology, and these perceptions may be influenced by situational cues. After all, relationships are part of our socially constructed reality (Berger & Luckmann, 1966) that is enacted through symbolic performance (Goffman, 1959). Power and status, for example, are ceremonially enacted across cultures (Fiske, 2004), and simply granting someone the trappings of power can make them act (Chen, Lee-Chai, & Bargh, 2001) and be perceived (Schwartz, Tesser, & Powell, 1982) accordingly. Furthermore, the influence of situational cues on relationship perceptions are important because many relationships are characterized by more than one relational model, and it is the perception of which model best describes a particular situation that guides actors’ behavior. For example, I could perceive my relationship with my colleague as communal
sharing when it comes to lending a book, authority ranked when we work on a project in which she is the lead, and market priced when she sells me her car. Each situation has different cues which influence my perception of the relationship and thus my motivations in that interaction. Importantly, if the cues present in a situation reduce the likelihood of perceiving an authority figure, then it may become more difficult to morally disengage, and hence compliance with unethical instructions may be reduced. Consistent with this argument, Milgram (1974) found that when the unethical orders came not from a Yale researcher wearing a lab coat, but rather from a plain-clothed member of an unknown organization, “Research Associates of Bridgeport,” compliance declined significantly.

I predict that pay-for-performance will reduce compliance with unethical instructions because the employer-employee relationship is inherently ambiguous and pay-for-performance serves as a cue that is inconsistent with an authority ranked relationship. Employees can view their relationship with their employer as either an authority ranked relationship, governed by a hierarchy in which they take orders from their superiors, or as a market priced relationship in which they sell services to their employer for remuneration. These two basic conceptions of the employee-employer relationship have been previously described as mutual-investment and quasi spot-contract relationships (Tsui, Pearce, Porter, & Hite, 1995; Tsui, Pearce, Porter, & Tripoli, 1997), respectively, and they also map onto theoretical distinctions between social and economic exchanges (Blau, 1964; Heyman & Ariely, 2004) or normative and utilitarian involvement (Etzioni, 1961).

Pay-for-performance should increase employees’ tendency to perceive their employment relationship as market priced for three reasons. First, pay-for-performance resembles a market exchange of service for money because it is structured according to the moral principle of a
Market priced relationships are governed by the principle of proportionality measured by a common medium of exchange, which is typically money, (Rai & Fiske, 2011) and in pay-for-performance remuneration structures employees receive payments that are proportional to their outputs. Second, pay-for-performance makes it easier for employees to calculate the monetary value of each action they perform in their job, and thus should make money more salient while they work. As money is the most ubiquitous symbol of market pricing (Fiske, 2004), its involvement in exchanges has been found to induce congruent relational construals (Gneezy & Rustichini, 2000; Heyman & Ariely, 2004). Third, performance-based pay acts as a financial incentive for compliant behavior, and as such signals the absence of normative pressure to comply (Grant, 2012). In essence, such an incentive signals that the employer lacks confidence in their authority to motivate compliance without the incentive. As a result, financial incentives often have “hidden costs” because they strip the principal-agent relationship of its normative force (Falk & Kosfeld, 2006; Fehr & Falk, 2002; Fehr & List, 2004). Thus, because pay-for-performance is structured according to proportionality, increases the salience of money, and signals the absence of authority, it should encourage employees to perceive their employment relationship as market priced rather than authority ranked.

Holding a market-priced perception of the employment relationship, as opposed to an authority ranked perception, should decrease the tendency to morally disengage in the face of unethical instructions and will therefore decrease compliant behavior. By deploying salient financial incentives to induce compliance an employer signals insufficient authority to compel compliance. Without that authority, an employee is less likely to perceive a moral obligation to obey and more likely to put forth effort they deem proportional to the incentive. Of course if the incentive is sufficiently large it will encourage proportional destructive conformity, but this may
ultimately prove inefficient because pay-for-performance also hinders moral disengagement. Not only do financial incentives undermine the perception of an authority on which to displace moral responsibility, they also exert an individuating effect by priming money. Multiple studies have consistently found that because money is both a tool and a symbol of self-advancement, merely making money more salient activates an individualistic mindset that reduces concern for others (e.g., Reutner & Wanke, 2012; Vohs et al., 2006; Vohs, Mead, & Goode, 2008). Thus although money primes can sometimes increase unethical behavior (Kouchaki, Smith-Crowe, Brief, & Sousa, 2013), the salience of money may actually impede compliance with unethical instructions because they increase an employee’s sense of independence. Individuals, acting for their own monetary gain, are responsible for their own actions and thus less likely to behave unethically because of the negative consequences for their moral self-regard. Thus, rather than feeling obligated to obey an employer who seems responsible for employees’ behavior, pay-for-performance should enhance employees’ perception of a market priced relationship in which individually responsible actors are engaging in voluntary exchange for their own benefit and thereby reduce compliance with unethical instructions. Therefore, I make the following predictions, graphically represented in Figure 1, which I will test with two experiments:

*Hypothesis 1.* Pay-for-performance will reduce compliance with unethical instructions compared to flat-fee incentive structures with the same expected value.

*Hypothesis 2.* This effect will be moderated by the employee’s ethical evaluation of the request, such that pay-for-performance will increase compliance with instructions the employee views as amoral, or ethically neutral, as previous studies have observed.
Hypothesis 3. The effect of pay-for-performance on compliance with instructions will be mediated by employee’s perception of the employee-employer relationship as authority ranked.

Hypothesis 4. The effect of pay-for-performance on compliance with instructions will also be mediated by employee’s perception of personal responsibility for their actions.

--------------------------------
Insert Figure 1 about here
--------------------------------

**Methodology Overview**

In order to test these hypotheses I am proposing two experimental studies. The experimental method is the most appropriate approach for testing these hypotheses not only because it can assess my causal arguments, but also because attraction, selection, and attrition processes (Schneider, 1987) are likely to produce range restriction in field samples. If, as I propose, pay-for-performance decreases employees’ ability to morally disengage and carry out tasks they find unethical, then it may also influence which employees are most likely to be found doing dirty work. Individual differences in the extent to which an employee values moral self-regard (Aquino & Reed, 2002) or libertarian moral values (Iyer, Koleva, Graham, Ditto, & Haidt, 2012) may enable certain employees to engage in destructive conformity without the need to morally disengage. Thus, to avoid sorting effects I have chosen to assign participants to unethical tasks in experiments which should allow me to observe a wider range of behaviour and thus a better chance of detecting the proposed phenomenon. While it is true that laboratory experiments place limits on external validity, gaining a better understanding of the
psychological phenomenon underlying destructive conformity and moral disengagement in the general population would still represent a significant theoretical contribution.

As a means of identifying the best methodology for testing my hypotheses, I explored three different experimental paradigms with pilot tests, which I describe below. Two of these experimental paradigms, a modified ultimatum game (Boles, Croson, & Murnighan, 2000) and bug grinding task (Martens, Kosloff, Greenberg, Landau, & Schmader, 2007), were developed by other scholars, and the third paradigm, a tobacco marketing task, was developed for this dissertation. On the basis of results from these pilots I have chosen to test my hypotheses with the bug grinding and tobacco marketing paradigms and have made some additional modifications to these paradigms in response to the results of Pilot Tests 2 and 3. Experiments 1 and 2 therefore serve as the ultimate tests of my hypotheses for the purposes of this dissertation.

**Pilot Test 1**

In Pilot Test 1 a laboratory experiment was conducted using a modified version of the ultimatum game (Boles et al., 2000). A standard ultimatum game consists of an economic transaction between two players assigned to divide a monetary endowment between themselves (see Roth, 1995, for a review). One player is assigned the role of ‘proposer’ and makes the initial decision of how to split the money. The second player is the ‘responder’ and decides whether to accept or reject the proposal. If the proposal is accepted, the money is divided between the players accordingly, but if it is rejected both players receive nothing. From the perspective of rational actor models, responders should accept any split in which they receive more than zero dollars, but in reality they frequently reject proposals that would give them less than half of the money. In anticipation of this, proposers tend to offer close to half of the endowment in an attempt to avoid a rejection (e.g., Oosterbeek, Sloof, & van de Kuilen, 2004).
The two modifications made to the ultimatum game in this experiment were first, the presence of an agent who made decisions on behalf of the proposer, and second, asymmetrical information regarding the sum of money being divided. The nature of this information asymmetry, the proposer and their agent knew the exact value of the monetary endowment whereas the responder was only told that this amount could range between $10-$40, created the possibility of deception (Boles et al., 2000). Specifically, agents could lie about the size of the monetary endowment to increase the likelihood that small offers would be accepted by responders, and in this study they received instructions from a greedy proposer that encouraged them to lie.

**Participants and Procedure**

**Participants.** One-hundred fifteen undergraduate students enrolled in introductory courses at a large North American business school participated in “A Computer Based Decision-Making Study” for course credit. This sample was 58.3 percent female, ranged in age between eighteen and twenty-four years ($M = 20.5$, $SD = 1.4$), and had an ethnic composition that was 57.4 percent East Asian, 16.5 percent White, 13.9 percent South Asian, and 12.2 percent other or unspecified.

**Procedure.** Participants were brought into a large computer lab with 18 terminals divided by privacy screens. Once participants had signed the consent form, a researcher read an announcement in which participants were told that they would be engaging in an anonymous monetary decisions task with other people in the room over the computer network. Next participants read instructions for the monetary decision task at their own pace on the computer screen. These explained that each participant would be randomly assigned to one of three roles, proposer, agent, or responder, for all six rounds of a monetary decision task. Although proposers
interacted with a different responder in each round, proposer-agent pairs stayed together for the duration of the study. As described above, the monetary decision task was a modified ultimatum game in which the proposer and responder were tasked with dividing a different endowment of money each round, one of which would be selected at random to serve as proposer and responders’ payout at the end of the study. Importantly, although the proposer and their agent knew the exact amount of money to be divided each round, the responder only knew that the amount varied between $10-$40. Thus, each proposal consisted of a declaration of the total endowment amount and the amount that the responder could receive if they accepted the proposal. It was also explained to participants that the agent would make proposals on behalf of the proposer for at least half of the 6 rounds. The agent was automatically assigned the job of making proposals in the first half of the study, but the proposer had the option of making the proposals themselves in the second half or “rehiring” the agent to continue making the proposals for the last three rounds.

Unbeknownst to participants, the monetary decision task was actually a computer simulation in which all of the participants interacted solely with a computer program. All participants were assigned the role of proposer or agent. Approximately one third of participants engaged in a two-player modified ultimatum game as a proposer with no mention of the agent role (otherwise this game had the same structure as the game described above). The remainder of participants were assigned the role of agent and told that they would be making proposals on behalf of another participant who was their proposer. Agents were further told that to help them make the proposal decisions they would receive information from both their proposer and the responders regarding what both expected to receive in each round. At the beginning of this round this information was presented in the form of the dollar amount that the proposer wanted
to keep for themselves from the endowment and the minimum percentage of the total endowment that the responder would be willing to accept. The expectations of the proposer and responder were always such that the agent had to misrepresent the total endowment amount in the proposal in order to reconcile these two expectations. Agents were led to believe that meeting their proposers’ expectations was in their own interest because the proposer had the option of making the proposal decisions without the agent in the second half of the monetary decision task (rounds 4, 5, and 6). Importantly, continuing to act as the agent in the second half of the monetary decision task would mean that the agent would receive more remuneration for their work. Half of agents were compensated via a flat pay structure and received $7 for each half of the monetary decision task in which they made proposal decisions (thus these agents believed that they could earn either $7 or $14 depending on whether they were hired back by the proposer for the second half of the monetary decision task). The other half of participants received performance-based pay, and received one-sixth (16.7 percent commission) of their proposers potential payout for each round in which they made the proposal decisions. At the end of the study, participants were debriefed and then paid the outcome of the monetary decision task in private before being dismissed.¹

**Manipulation and Measures**

**Manipulation.** The computer program randomly assigned each participant to one of three conditions. Participants assigned to the role of proposer played the modified ultimatum game computer program with no agent and were informed that the outcome of one of the rounds would be chosen at random for their payment at the end of the study. Participants assigned to the

¹ To view the actual computer program used for this pilot test, visit the following website: http://rotman.az1.qualtrics.com/SE/?SID=SV_cISLFMrrYfnsniZ
role of agent with a flat pay structure were told that they would automatically earn $7 for making proposal decisions on behalf of their proposer during the first half of the study and that there was the possibility of earning another $7 if they were “rehired” to make proposal decisions for their proposer in the second half of the study. Finally, participants assigned to the role of agent with a performance-based pay structure were told that they would be paid one-sixth of their proposer’s potential earnings for each round in which they made the proposal decisions on behalf of their proposer, and that although they were automatically assigned to make these decisions in the first half of the study, they would only make decisions in the second half of the study if they were “rehired.” In actuality, all participants were “rehired” by their proposer for the second half of the study.

**Dependent variables.**

*Underreporting the monetary endowment.* After participants had read the monetary decision task instructions and correctly answered six questions which tested their comprehension of the task, participants completed a practice round to familiarize themselves with the user interface. At this point participants were “randomly assigned” a proposer and responder for the first round of the monetary decision task (those participants playing the role of proposer were only assigned a responder), and received information about the monetary endowment for that round as well as these other players’ expected outcomes after a slight time delay designed to increase the realism of the interaction. The monetary endowment in each round varied between $10 to $40 in such a way as to give the impression of randomness (the endowments for each round were as follows, round 1: $28, round 2: $18, round 3: $34, round 4: $38, round 5: $22, round 6: $20). In each round the responder would indicate that they expected to be offered between 40 to 50 percent of the monetary endowment, and for participants playing
the role of agent their proposer would declare expectations of between 72 to 79 percent of the endowment. Thus in every round, if participants assigned to the role of agent wanted to satisfy the expectations of their proposer, they had to lie to the responder and underreport the true value of the monetary endowment so it would appear that they were offering the responder a proposal that met responder’s demands. The responder was programmed to reject proposals that fell below their declared expectation as well as those proposals in which the participant declared a total monetary endowment less than $10 (responders knew that the endowment amount varied between $10-$40). Therefore, as the magnitude of underreporting in which participants could engage during each round varied as a function of each round’s monetary endowment, the focal dependent variable was the proportion of maximal underreporting in which the participant engaged, calculated according to the formula: $\frac{E-D}{E-10}$ where $E$ is the total monetary endowment to be split between the proposer and responder in the round and $D$ is the monetary endowment declared by the participant in the proposal. At the end of each round participants were reminded of their pay structure and given an updated record of their pay outcomes for each previous round, and when the monetary decision task had finished each participant was shown their total pay.

**Donation opportunity.** Immediately after finishing the monetary decision task, participants were presented with the opportunity to make an anonymous donation to the SickKids Hospital Foundation. Participants were reminded of the amount that they would be paid at the end of the study and were told that they could donate any part of this amount to the charity and that this donation would be automatically deducted from what they would be paid. Providing participants with this opportunity was intended as a behavioural measure of guilt. When people’s positive moral self-concept is threatened, they often engage in compensatory
ethical behaviours in order to restore their moral self-concept (Zhong, Ku, Lount, & Murnighan, 2010), such as donating to charity (Sachdeva, Iliev, & Medin, 2009).

Mediating variables.

**Affect.** Although I have argued for a more cognitive mechanism of relationship construal and assessment of personal responsibility, it is also possible that the behavioural effect of pay structure on unethical behaviour could be carried by a more diffuse affective pathway. For that reason, after deciding how the donation task, both state positive and negative affect were measured by the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). The PANAS includes 20 single-word items which describe a wide range of positive and negative emotional states. Participants were asked to indicate the extent to which each word described how they were feeling at the present moment (1, “very slightly or not at all,” to 5, “extremely”). Reliability estimates for the positive, $\alpha = .90$, and negative, $\alpha = .87$, subscales of this measure were high so an average score for each was calculated.

**Perceived authority.** To test Hypothesis 3, I sought to measure participants’ perception that their relationship with their proposer was authority ranked. All participants were therefore presented with two Likert-type items randomly intermixed with the other mediating measures discussed below. These items asked participants to what extent they agreed or disagreed with the statements “It was my job to obey the instructions I received in the decision-making task” and “I was in control of the proposal decisions in the task” (1, “strongly disagree” to 7, “strongly agree”). Although the second item was intended to be reverse coded, the correlation between the items was negative after reverse coding, $r = -.244$, $p = .009$, so these items were not averaged together to form an index of authority ranked relationship perceptions.
**Perceived personal responsibility.** To test Hypothesis 4, I sought to measure participants’ perception that they were personally responsible for the outcomes of the monetary decision task. Two Likert-type items assessed personal responsibility perceptions: “I felt responsible for the content of the proposals in the decision-making task” and “I should receive the credit or the blame for the outcomes of the decision-making task,” and both used the same scale as the authority ranked perception items. These two items were significantly and positively correlated, $r = .524, p < .001$ and were thus averaged into an index of perceived personal responsibility.

**Perceived personal benefit.** Another possible mechanism which could explain the effect of pay structure on individuals’ willingness to comply with unethical instructions is perceived personal benefit. If someone benefits from unethical conduct others perceive them as morally responsible for that conduct, even if they did not cause or contribute to its occurrence (Inbar, Pizarro, & Cushman, 2012). In a similar way, perceiving one’s own personal benefit from unethical conduct might make an individual feel a greater sense of responsibility for that conduct and therefore less likely to behave unethically. Conversely, people are more likely to cheat when third parties share in the benefit because this makes the act of cheating seem less unethical (Wiltermuth, 2011). When employees receive pay-for-performance, their personal benefit for complying with instructions is clear and calculable. In contrast, flat pay structures obscure employees’ personal benefit because they get paid the same irrespective of the level of their compliance with instructions, at least in the short term. Thus, this increased transparency in personal benefit from unethical behaviour may reduce compliance with unethical instructions under pay-for-performance. Two items, “I benefited personally from the proposal decisions in

---

24
the task” and “I made decisions on the basis of my own interests in the task” were developed to measure this perception and averaged due to their high correlation, $r = .53$, $p < .001$.

**Results**

Overall the results from Pilot Test 1 were disappointing. Contrary to Hypothesis 1, a mixed-method three (no agent by commission agent by flat fee agent) by six (modified ultimatum game rounds) analysis of variance with repeated measures on the second factor revealed the opposite pattern from that predicted on participants’ proportion of maximal lying. Although there was an overall between participants effect, $F(2, 112) = 8.29$, $p < .001$, this was driven by the fact that participants assigned to play the role of agent lied to a greater extent than those who played the modified ultimatum game representing only themselves. This finding is consistent with previous research showing that people tend to act more selfishly when representing someone else as opposed to themselves (Hamman, Loewenstein, & Weber, 2010). Importantly, however, while a planned simple contrast did not find a significant difference between the effect of pay structures on the size of agents’ lies, $p = .235$, the pattern of means was opposite to the expected pattern, as can be seen in Figure 2. The pay structure of agents also did not have an effect on their donation behavior either with or without controlling for the amount that they were paid, $b(112) = .055$, $p = .61$. Finally, planned contrasts revealed that pay structure did not have a significant effect on the positive or negative affect, relational construal, or perceived personal responsibility, all $F$-statistics $< 1$. The only finding which I was able to predict was that performance-based pay did increase the agents’ perception of personal benefit ($M = 5.18$, $SD = 1.39$) relative to flat pay structures ($M = 4.31$, $SD = 1.40$), $F(2, 112) = 1.72$, $p = .007$, although clearly this effect did not mediate any behavioural changes measured in this experiment.
Due to these unpromising results, I have decided to pursue other experimental paradigms to test my hypotheses. Ultimately null results are impossible to interpret conclusively, but perhaps this paradigm is ill suited to study the proposed processes because of its high-levels of anonymity; all actors in this modified ultimatum game remained anonymous and unidentified. Previous research has found that public visibility has a moderating effect on the crowding-out effect of monetary incentives on prosocial behavior (Ariely, Bracha, & Meier, 2009). Thus it could be that agents’ privacy interfered with the effect of pay structure on their willingness to behave unethically on behalf of their principals. The possibility of public scrutiny may strengthen agents’ concern for their moral self-regard as an estimate for how they are likely to be judged by others. An alternative explanation for these null results could be that an unidentified, anonymous principal constitutes a poor authority figure on which to displace moral responsibility. The excuse of simply “following orders” may only apply if those orders have the moral backing of a legitimate authority figure, and it is likely impossible to determine an authority’s legitimacy when the authority’s identity is unknown. For these reasons, the other pilot tests I designed feature an identifiable authority figure and in Pilot Test 3 participants were observed as they engaged in the unethical act.

--------------------------------

Insert Figure 2 about here

--------------------------------

**Pilot Test 2**

Pilot Test 2 involves a novel experimental paradigm inspired by the quasi-legal and increasingly popular business practice of online tobacco advertising (Campaign for Tobacco-Free Kids, 2011). In light of the heavy regulation on tobacco advertising in broadcast and print
media, tobacco companies have turned to advertising online because this has yet to be regulated by the Federal Communications Commission and it provides ready access to younger consumers. Participants recruited for this pilot test were invited to produce advertising material for use in online social media. Participants were instructed to write advertising messages that promote social smoking among young consumers and told that these messages would be evaluated and potentially used by Philip Morris in an advertising campaign. Participants were paid for their efforts, purportedly from funding that had been provided by Philip Morris, and the structure of this payment either took the form of an unconditional payment or payment based on the amount of advertising material they produce during the study.

**Participants and Procedure**

**Participants.** 123 students enrolled in introductory management courses at a large North American university were recruited for an online “Creative Marketing Ideas Study” and offered course credit as compensation for their time. This sample was 61.3 percent female, ranged in age between eighteen and twenty-four years ($M = 20.3$, $SD = 1.5$), and had an ethnic composition that was 48.1 percent East Asian, 21.7 percent White, 17.9 percent South Asian, and 13.2 percent other or unspecified. Sixteen participants were excluded from analyses because they reported technical difficulties while completing the online study.

**Procedure.** The cover story for the online experiment was that a group of marketing researchers at the business school operate a “Creative Marketing Ideas” institute which conducts studies with industry partners. Participants were told that industry partners approach the researchers with advertising tasks with which they want the help of young creative minds, and that this arrangement allows the researchers to study creative problem solving on real-world advertising tasks. Furthermore, in exchange for funding the research, which included incentive
payments to participants to make the problem solving more similar to conditions in the advertising industry, it was explained that industry partners were entitled to use the advertising content generated in the studies for their advertising campaigns. At this point participants were shown the logos of several previous industry partners and then informed that the sponsoring partner for the study in which they were participating was Philip Morris International. A brief description of Philip Morris’s corporate history, status as the world’s largest tobacco company, and its highly successful Marlboro brand, proceeded instructions for the advertising task with which the company supposedly wanted help. Participants read that although Philip Morris continued to see strong growth in many developing economies, sales were flagging in North America, especially among young consumers, and that market research had indicated this was the result of cigarettes no longer being seen as ‘cool’ or ‘sexy’ on account of their potential link to negative health consequences and addiction. To address this situation Philip Morris was planning a social media advertising campaign to be conducted on the micro-blogging service Twitter and expressed excitement about partnering with the Creative Marketing Ideas Institute because its marketing department hoped that the student participants could help the company find the youthful voice necessary to make the campaign a success. Participants were provided with examples of the kinds of advertising messages that would be appropriate and reminded of the monetary payments they would receive for participating in the study.

On the following page participants were invited to write up to 12, 140 character advertising messages for Philip Morris. Participants were randomly assigned to either receive a flat payment of $10 for this work, irrespective of their performance, or to be paid $0.85 per advertising message they composed. Both groups of participants were instructed to choose one of several appointments in the future when they could receive their cash payments in private.
This delay in payment was designed to give participants the impression that their advertising messages would be read and evaluated before they received their compensation. After completing this advertising task, participants filled out a number of questionnaires and were provided the opportunity to donate their earnings in the study to a cancer research charity before being debriefed about the true nature of the study.²

**Manipulation and Measures**

**Manipulation of pay structure.** Participants were randomly assigned to receive either a flat payment of $10 or pay-for-performance in the form of $0.85 per advertising message they produced, up to a total of 12 messages. These two payment structures have approximately equal expected value, in that participants in the pay-for-performance condition could earn a maximum of $10.20 if they produce 12 advertising messages. Participants were informed about this pay structure twice, once at the beginning of the study when they selected a future pay-day appointment before they learned that Philip Morris International was the industry partner for this study, and then a second time immediately before they completed the advertising task.

**Dependent variables.**

**Task performance.** Participants’ performance on the advertising task was measured by counting the number of supportive advertising messages (i.e. messages that were explicitly anti-tobacco were not counted as performance of the task).

**Task effort.** The amount of time for which participants worked on the advertising task was recorded and used as a measure of effort on the task.

²To view the actual computer program used for this pilot test, visit the following website: [http://rotman.az1.qualtrics.com/SE/?SID=SV_4JaPV7ufsbl9gs5](http://rotman.az1.qualtrics.com/SE/?SID=SV_4JaPV7ufsbl9gs5)
Donation opportunity. After completing all of the questionnaires and demographic measures participants were presented with the opportunity to make an anonymous donation to the Canadian Cancer Society. Participants were reminded of the amount that they would be paid at the end of the study and were told that they could donate any percentage of this amount to the charity and that this donation would be automatically deducted from what they would be paid.

Mediating variables.

Perceived authority. Three items were designed to measure the extent to which during the advertising task participants perceived their relationship with Philip Morris as authority ranked. The questions “to whom did you feel it was your job to obey while writing the messages?,” “who do you feel was more in charge of what the messages you wrote were about?,” and “Did you feel an obligation to yourself or to Philip Morris International when writing the marketing messages?” were presented with 7-point Likert-type scales that were tailored to each statement (e.g., 1, “obligation to obey only myself,” to 7, “obligation to obey Philip Morris International”). These three items exhibited high reliability, α = .79, and thus were averaged together to create a perceived authority index.

Perceived personal responsibility. Two items were designed to measure the extent to which participants perceived themselves or Philip Morris to be responsible for the content of the advertising messages that they wrote. The questions “who do you feel is more responsible for the content of the messages you just wrote on Philip Morris’s behalf?” and “who do you feel should receive the credit or blame for the outcomes of the messages you wrote?” were presented with 7-point Likert-type scales tailored to each item (e.g., 1, “I am entirely responsible,” to 7, “Philip Morris International is entirely responsible”). As these two items were highly and
significantly correlated, \( r(104) = .595, p < .001 \), they were averaged together and reverse scored to create a perceived personal responsibility index.

**Perceived personal benefit.** As in Pilot Test 1, two items were used to measure the alternative mechanism of perceived personal benefit. These items were “who benefited more from the messages that you wrote?” and “In whose interests were you acting when you wrote the messages on the previous page?” and they were measured on 7-point Likert-type scales tailored to each item (e.g., 1, “Only I benefited,” to 7, “Only Philip Morris International benefited”). As expected, these items were positively and significantly correlated, \( r(104) = .59 \), \( p < .001 \) and therefore averaged and reverse scored to create an index of perceived personal benefit.

**Affect.** As in Pilot Test 1, an alternative affective mechanism was measured, although unlike in Pilot Test 1 this measurement took the form of a semantic differential scale (Mehrabian & Russell, 1974). Embedded within other items of bipolar adjective pairs, were two items that measured the moral emotion of guilt. Participants were asked to indicate on a 7-point scale anchored between the adjectives “proud – ashamed” and “guiltless – guilty” how they felt at the present moment. The positive and significant correlation between these two items, \( r(105) = .73, p < .001 \), indicated that it was appropriate to average them into an experienced guilt index.

**Moderating variables.**

**Moralization of tobacco advertising.** For the theoretical mechanism proposed in this paper to apply, an employee must subscribe to the hypernorm he or she is being asked to violate. Otherwise they will not view the request as immoral and will not experience any threat to their moral identity and a subsequent need to morally disengage if they are to comply with the
request. Therefore, because there may be some variance in the extent to which participants drawn from the student population view tobacco advertising that targets young consumers as unethical I sought to measure this attitude. After participants had completed the aforementioned mechanism measures they were presented with a 7-point semantic differential scale consisting of adjectives that described Philip Morris International. Four bipolar adjective pairs were intended to index participants’ ethical attitude towards the company: “good – bad,” “desirable – undesirable,” “moral – immoral,” and “helpful – harmful.” The internal reliability of these four items was high, $\alpha = .93$, and so they were averaged and reverse scored to create an index of participants’ ethical evaluation of Philip Morris International, the purported employer in this study. Immediately after these items participants answered the question “How moral or immoral (ethically good or bad) do you consider to be the advertising of tobacco products to consumers aged 19-25?” on a 7-point Likert-type scale (1, “highly moral,” to 7, “highly immoral”).

**Personal use of tobacco products.** It is possible that whether or not one personally uses tobacco products may shape one’s attitude toward tobacco advertising. Therefore, once participants had completed the advertising task, they were asked to indicate the frequency of their personal use of tobacco products on a 7-point scale (1, “Never,” to 7, “Daily”).

**Moral identity centrality.** An assumption of my model, which draws on Mazar, Amir, and Ariely’s (2008) self-concept maintenance theory, is that people are motivated to preserve their moral self-concept. Research by Aquino and Reed (2002), however, suggests that people vary in the extent to which they treat their moral self-concept as central to their identity. People for whom moral self-concept is a relatively unimportant component of their identity are unlikely to feel the same aversion to acting contrary to society’s hypernorms and therefore may also feel less need to morally disengage when presented with opportunities to benefit by violating those
norms. For exploratory purposes, I therefore administered the Aquino and Reed’s moral identity scale (2002) after participants had answered all of the other items about their ethical evaluation of the advertising task, although I had no a priori hypotheses regarding these variables. Both the internal and symbolic moral identity subscales exhibited high reliability, $\alpha = .84$ and $\alpha = .75$, respectively.

**Suspicion probe.** Immediately after the advertising task participants were asked two open-ended questions that were designed to surreptitiously probe the extent to which they believed the study’s cover story. These questions were “How much cash do you expect to receive on the payday you chose to attend next week?” and “Who funded this study?” Participants responses to these questions were coded for any indication that they did not believe that they would be paid or that the money they were to receive did not come from Philip Morris International. The answers of nine participants indicated some level of disbelief in either their payment for the advertising task or Philip Morris’s sponsorship of the study and thus these participants were removed from the analyses.

--------------------------------
Insert Table 1 about here
--------------------------------

**Results**

Bivariate correlations between all of the aforementioned variables from Pilot Test 2 are displayed in Table 1. As a test of Hypotheses 1 and 2 I conducted four multivariate linear regression analyses examining the relationship between participants’ performance on the advertising task and their assigned pay structure as moderated by their moralization of tobacco advertising and Philip Morris, their habitual smoking behaviour, and their moral identities.
Although the regression models featuring the length of time participants’ spent on the advertising task, their moral attitude towards Philip Morris and their habitual smoking behaviour did not show any significant relationships, the model which included the number of advertising messages generated and participants’ moralization of tobacco advertising, the specific behaviour for which participants were paid in this study, did uncover the expected pattern of relationships. Specifically, although there was no main effect of participants’ moralization of tobacco advertising, $\beta = -.03, t(96) = -0.21, p = .833$, pay-for-performance significantly increased the number of messages participants produced as agency theory would predict, $\beta = .829, t(96) = 2.22, p = .031$. Importantly this main effect of pay for performance was qualified by a marginally significant interaction with participants’ moral attitude towards tobacco advertising, $\beta = -.76, t(96) = -1.87, p = .064$, such that pay-for-performance was less effective at motivating the production of advertising messages for those participants who judged tobacco advertising as more unethical compared to those who viewed tobacco advertising as morally acceptable. Figure 3 presents the nature of this relationship graphically, following the procedure outlined by Aiken and West (1991). Simple slopes analyses confirm that for those participants who moralized tobacco advertising to a lesser extent (1 standard deviation below the mean moralization in the sample) performance-based pay had a significant positive effect on their performance relative to a flat pay structure, $\beta = 2.82, t(96) = 2.40, p = .019$, whereas for

---

3 Length of time spent on the advertising task was also examined with a log transformation to correct for its significant positive skew.

4 Moralization of tobacco advertising was also analyzed as a dichotomous variable, such that those participants who viewed tobacco advertising as slightly, moderately or very immoral, were contrasted with those who viewed tobacco advertising as amoral or virtuous. The same pattern of results were observed, a significant main effect of performance-based pay on performance, $\beta = .41, t(96) = 2.40, p = .018$, which was qualified by a significant interaction with the dichotomized moralization of tobacco advertising, $\beta = -.40, t(96) = -2.00, p = .048$. 34
those who moralized tobacco advertising to a greater extent (1 standard deviation above the mean) pay-for-performance had no greater motivating effect than did flat pay structure, $\beta = -.29$, $t(96) = -0.25$, $p = .800$. It should also be noted that as a robustness check participants’ habitual smoking behaviour was controlled for, because smoking is related to moral attitudes towards tobacco advertising and might conceivably make the generation of creative tobacco advertising messages easier due to greater familiarity, and the relationships were unchanged. A similar pattern of results was observed with participants’ internal, but not symbolic, moral identities, such that although internal moral identity did not have a significant main effect on performance, $\beta = .16$, $t(96) = 1.16$, $p = .250$, the significantly positive main effect of pay-for-performance, $\beta = 1.37$, $t(96) = 2.07$, $p = .042$, was moderated by participants’ internal moral identity centrality, $\beta = -.129$, $t(96) = -1.9$, $p = .060$ (see Figure 4 for the pattern of results). In contrast there was no effect of pay structure, the moralization of tobacco advertising, or their interaction on participants’ donation behaviour, all $t$-values < 1.

In order to test whether the interaction between pay structure and moralization of tobacco advertising in predicting performance on the advertising task could be explained by any of the proposed mediators I conducted regression analyses predicting perceived authority, perceived personal responsibility, perceived personal benefit, and self-reported guilt. Of all of the proposed mediators, however, only perceptions of authority exhibited a significant regression coefficients. Specifically I found that although participants’ moralization of tobacco advertising was not significantly related to whether they perceived their relationship with Philip
Morris as authority ranked, $\beta = .186$, $t(90) = 1.28$, $p = .203$, pay-for-performance exerted an overall positive effect on participants’ perception of Philip Morris’s authority in the advertising task, $\beta = .95$, $t(90) = 2.28$, $p = .025$. As with participants’ performance on the advertising task, however, this relationship was qualified by a significant interaction with moralization of tobacco advertising, $\beta = -1.09$, $t(90) = -2.47$, $p = .015$, and the nature of this interaction can be seen in Figure 5. Simple slopes tests indicate that although authority perceptions of participants who moralized tobacco advertising to a lesser extent were not affected by performance-based pay, $\beta = .77$, $t(96) = 1.49$, $p = .139$, those participants who more greatly moralized tobacco advertising perceived their relationship with Philip Morris as less authority ranked when they received performance-based pay as compared to flat pay, $\beta = -1.04$, $t(96) = -2.05$, $p = .043$. To test the mediated moderation model I propose (without the second mediator of personal responsibility due to its insignificant relationships with the independent and moderating variables) I conducted a bootstrapped conditional indirect effect test as per Preacher, Rucker, and Hayes, (2007) using model number 8 of the PROCESS macro for SPSS (Hayes, 2013). Based on bootstrapped samples of 10,000, the bias corrected 95% confidence interval for the indirect effect of pay-for-performance on performance in the advertising task through perceptions of an authority ranked relationship, conditional upon moralization of tobacco advertising did not cross zero $[-1.00, -0.02]$. This is significant evidence of an indirect effect suggesting that part of the reason why performance-based pay is less effective at motivating unethical work is because it reduces participants’ perception that their relationship with an employer is authority ranked.

-----------------------------

Insert Figure 5 about here

-----------------------------
Discussion

On the whole Pilot Test 2’s results were more supportive of my hypotheses. As agency theory would predict, pay-for-performance did have a significant positive effect on participants’ performance on the advertising task. In support of Hypothesis 2, however, I found that this effect on performance was moderated by the extent to which participants moralized the advertising of tobacco products. That is performance-based pay only outperformed the flat pay structure among those participants who moralized tobacco advertising to a lesser extent. For those participants who judged tobacco advertising as an unethical act, performance-based pay was no more motivating than was a flat pay structure. Contrary to Hypothesis 1, however, I did not find that participants who viewed tobacco advertising as immoral performed worse under performance-based pay structures compared to flat pay structures. Importantly, I also found that this interaction could be at least partially explained by the effect of performance-based pay on authority perceptions. For those participants that viewed tobacco advertising as immoral, pay-for-performance, relative to a flat pay structure, reduced the extent to which they perceived their relationship with Philip Morris to be authority ranked, and this effect mediated the aforementioned interaction. Thus Pilot Test 2 provides some support for a simplified version of my model in which the mediating mechanism is composed of only authority ranked relationship perceptions. Given the relatively small sample size of the pilot test, however, it is unclear to what extent the partial support for my model is a product of a low powered study. For that reason, in light of this partial support I propose to expand Pilot Test 2 into a full scale study which will be able to test my model more definitively.

There are two major limitations to Pilot Test 2. The first and most critical limitation is that the measure of tobacco advertising moralization was taken after participants had completed
a task in which they were asked to engage in tobacco advertising. Therefore it could be that the experimental task influenced participants to judge tobacco advertising as more or less immoral as a means of justifying their recent behaviour. To address this limitation in Experiment 1 I measured the extent to which participants moralize tobacco advertising several weeks before they engage in the advertising task themselves. Another minor limitation of the current measure of tobacco advertising moralization is that as a single-item it is impossible to assess its reliability. Therefore I plan to use a multi-item measure of the extent to which participants moralize tobacco advertising in the proposed experiment. The other major limitation to Pilot Test 2 is that participants moralization of tobacco advertising was not manipulated. This reliance on natural variation in the extent to which participants view tobacco advertising as unethical helps to demonstrate a boundary condition of my model which applies in the real world, but it also leaves my findings open to third variable confounds. It is possible that people who differ in their views of tobacco advertising as either moral or immoral also differ across other individual differences, such as risk-aversion or libertarian values, which may also affect their response to performance-based pay (Cadsby, Song, & Tapon, 2007) or authority relationships (Iyer et al., 2012). To address this second limitation I tested a different experimental paradigm in which I planned to manipulate the extent to which the task participants were paid to do was immoral.

**Pilot Test 3**

Pilot Test 3 was thus designed to address this latter limitation by creating a paradigm in which participants could be assigned to an experimental task that was either ethically neutral or unethical and otherwise as similar as possible across conditions. The unethical task in Pilot Test 3 is adopted from Martens, Kosloff, Greenberg, Landau, and Schmader’s (2007) study of killing captive insects in the laboratory. Previous research has found that participants find executing
this behaviour to be moderately unethical and stress inducing (Webber, Schimel, Martens, Hayes, & Faucher, 2013; Wiltermuth, 2012). Wiltermuth (2012) argues that this paradigm serves as a conceptual replication of Milgram’s (1974) studies of destructive obedience, in which an authority figure instructs participants to engage in unethical behaviour. Indeed, similar to Milgram, Wiltermuth finds that actions which change participants’ construal of their relationship with the researcher influence their compliance with instructions to engage in purportedly lethal behaviour. In Milgram’s studies it was documented that compliance with unethical instructions was increased when these instructions came from a more legitimate authority figure, as opposed to a more dubious one. Wiltermuth (2012) shows that when participants marched in synchrony with the experimenter they were more likely to comply with the unethical instructions, as compared to when they marched out of step, apparently because the synchrony enhanced their sense of relational proximity to the experimenter. Thus this bug-killing task seems ideal for testing whether altering relational construals with pay structure might affect individuals’ willingness to comply with unethical instructions. In Pilot Test 3 I develop a novel adaptation of this paradigm which includes both the extermination of bugs and the same motoric behaviour applied to inanimate objects in a between-participants design. This allows me to test the same interaction between payment structure and unethical vs. ethically neutral behaviour that was observed in Pilot Test 2, while experimentally manipulating both factors.

**Participants and Procedure**

**Participants.** Forty-eight community members who had previously registered in a paid participant pool administered by a large North American business school volunteered for “A Study of Physiological Changes During Behaviour” in exchange for up to $10 in compensation.
This sample was 68.8 percent female, ranged in age between eighteen and forty-five years ($M = 23.1, SD = 5.2$), and had an ethnic composition that was 60.4 percent East Asian, 10.4 percent White, 18.8 percent South Asian, and 10.4 percent other or unspecified.

**Procedure.** Participants were brought into the lab individually. Those assigned to the flat-payment condition read on the consent form that they would be compensated $10 for their participation, irrespective of their behaviour in the study, whereas those assigned to the performance-based payment condition were informed that they would receive $5 irrespective of their behaviour in the study and up to an additional $5 depending on their performance in the experimental task. Otherwise the two versions of the consent form were identical. After participants indicated that they had read the consent form and had no unanswered questions the experimenter explained the cover story as they began to attach sensors to the participant’s non-dominant hand. The cover story was that the researchers were interested in people’s physiological responses during the execution of certain tasks, and once the physiological sensors were in place participants were taken to another room in which the experimental materials had been set up as displayed in Figure 6. For the experimental task participants were instructed to place as many of the provided objects as possible (up to 20) into a modified coffee grinder within 35 seconds and then turning the grinder on. In the ethically neutral control condition, the provided objects were 20 black beans which had been individually placed in plastic 1-ounce cups. In the unethical experimental condition the same cups were filled with one pill bug (*Armadillidium vulgare*) each and the modified coffee grinder was referred to as the “extermination machine.” Unbeknownst to participants, a plug in the modified coffee grinder prevented the beans and bugs from coming into contact with the grinding blades, and instead the grinding chamber was filled with shredded paper and beans to provide realistic auditory
feedback. After completing the experimental task participants returned to the room where they had read the consent form to answer a number of questionnaires. Finally participants were fully debriefed about the nature of the experiment, asked not to discuss the experiment with others for at least 3 months, and paid their compensation before being dismissed (see Appendix A for the laboratory protocol provided to research assistants).

Manipulation and Measures

Manipulation.

_Ethicality of experimental task._ Participants were randomly assigned to place either inanimate beans into a modified coffee grinder or living, captive bugs into an “extermination machine,” which they were lead to believe would result in the bugs’ death. Although these tasks are motorically identical, grinding beans is not expected to have any moral implications whereas previous research has found that participants find killing captive bugs moderately immoral and stress inducing (Webber et al., 2013; Wiltermuth, 2012). Both groups of participants were told by the experimenter that it was important that they put as many objects in the grinder as they could in the 35 second time limit so that the experimenter could obtain a reliable measure of their physiological response during this time. At the end of the 35 second period participants were instructed to push the grinders’ ‘on’ button, which activated the grinder for approximate 5 seconds and provided realistic auditory feedback. The experimenter was provided with a script that they were to read verbatim when explaining the grinding task, and this script included two
prompts designed to encourage compliance of hesitant participants before allowing participants to refuse to complete the experimental task.

**Pay Structure.** Participants randomly assigned to the flat-payment condition were told that they would receive $10 for their participation, irrespective of their performance on the experimental task. In contrast, participants randomly assigned to the pay-for-performance condition were informed that they would receive $5 irrespective of their performance, plus $0.25 for each object that they ground in the grinder. The expected utility of both conditions was exactly $10.

**Manipulation check.**

**Moralization of the experimental task.** To confirm that participants did view grinding living bugs as more ethically problematic than grinding beans they were asked the following question after completing the mediating items listed below. “How moral or immoral (ethically good or bad) do you think it was to obey the instructions of the researchers in this study?” (1, “Highly moral,” to 7, “Highly immoral”). Participants in the unethical task condition found the task significantly more immoral ($M = 4.96, SD = 1.27$) than participants in the control condition ($M = 2.37, SD = 1.28$), $t(46) = 7.03, p < .001$, and they also found the task significantly more immoral than the neutral scale midpoint, $t(23) = 3.70, p = .001$, indicating that the manipulation was successful.

**Dependent variables.**

**Task Performance.** The number of objects participants placed into the coffee grinder within the 35 second period was the primary dependent variable for this study.

**Voiced protest.** Employees’ willingness to voice concerns about their work is an important variable of interest to organizational scholars (Morrison, 2011), and since employees’
ethical leadership perceptions predict voice (Walumbwa & Schaubroeck, 2009), it would be valuable to know whether pay structure can mitigate the silencing effect of unethical leaders. Therefore the number of times participants refused to participate in the experimental task was measured as a secondary dependent variable. The maximum possible number of refusals was three, after which the experimenter was instructed to allow the participant to complete the questionnaires or withdraw from the study completely. Because only one participant refused to comply with the experimental instructions once there is insufficient variance to examine this variable further.

**Mediating Variables.**

**Perceived authority.** Three items from Pilot Test 2 were adapted to measure the extent to which during the experimental task participants perceived their relationship with the researchers as authority ranked. In contrast to Pilot Test 2 these three items exhibited questionable reliability, $\alpha = .60$, perhaps as a product of a relatively small sample size. These items were averaged together to create a perceived authority index.

**Perceived personal responsibility.** Two items from Pilot Test 2 were adapted to measure the extent to which participants perceived themselves or the researchers to be responsible for the consequences of their behaviour in the experiment. These two items were positively and significantly correlated, $r(47) = .35, p = .014$, and averaged together and reverse scored to create a perceived personal responsibility index.

**Perceived personal benefit.** Two items from Pilot Test 2 were adapted to measure the alternative mechanism of perceived personal benefit. These items were positively and significantly correlated, $r(47) = .50, p = .001$ and therefore averaged and reverse scored to create an index of perceived personal benefit.
**Affect.** A modified version of the PANAS (Watson et al., 1988) was employed in this study to measure guilt and disgust. Disgust was added to this study because it was thought that some participants’ disgust response to bugs may affect their performance on the experimental task. Three items intended to measure guilt, ‘remorse,’ ‘regret,’ and ‘guilt’ displayed excellent internal reliability, $\alpha = .93$, and were thus averaged together to create a self-reported guilt index. The two items intended to measure disgust, ‘grossed out’ and ‘disgust’ were also positively and significantly correlated, $r(47) = .80, p < .001$, and thus averaged into a self-reported disgust index.

**Suspicion probe.** At the end of the questionnaires participants in the unethical experimental condition were informed that they had not in reality killed any bugs using the extermination machine. At this point participants were asked to indicate whether they had previously known that the bugs were not really killed during the experiment. Of participants in the unethical task condition, 29.2 percent indicated that they had not believed that the bugs were actually killed in the experiment. These participants were therefore excluded from further analyses. Interestingly, it appears that participants were more likely to be skeptical of the “extermination machine” when they were in the flat payment condition (5 out of 10 participants expressed skepticism) compared to when they were assigned to receive performance-based pay (only 2 out of 14 participants in this condition expressed skepticism), $\chi^2 = 3.60, p = .058$, although I had no prediction nor explanation as to why this would be the case.

**Results**

Ultimately with a sample size of only 39 participants any conclusions are hard to draw. The associations among all of the previously described variables can be seen in Table 2. Although none of the hypothesized relationships are statistically significant with such a small
sample size I have taken two pieces of information from overall patterns produced by this pilot that will be useful when running a full-scale experiment (see Figure 7 for graphical representation of the results of Pilot Test 3). The first is that although previous researchers have found that a period of 40 seconds is an acceptable time for the bug grinding task, when it comes to the same task with beans 35 seconds provides participants too much time to complete the task. As a result, I believe that I may be observing a ceiling effect in my current findings. Thus, to increase the chances of finding a positive effect of performance-based pay relative to flat pay in the control condition I plan to shorten the amount of time participants have to complete the task to 30 seconds.

Second, my use of two research assistants to run Pilot Test 3 seems to suggest a potential boundary condition for my theory. One of the research assistants is a young female of small stature, while the other is male and of average height. Given the association between gender, height and perceived leadership and power (e.g., Gawley, Perks, & Curtis, 2009), it could be possible that participants are less likely to see the female research assistant as less of an authority figure, even though this difference does not appear in self-reported perceived authority index, $t(37) = .995, p = .326$. Nevertheless, because the presence of an authority figure issuing unethical instructions is central to my theory of destructive compliance, it seems wise to conduct this study with a research assistant who more closely resembles a stereotypical authority figure.

In the absence of a larger sample size I am reluctant to draw additional conclusions from this data. Overall the paradigm and measures appear to offer an acceptable test of my predictions and I therefore plan to put them to use with a larger sample size drawn from the same participant population.

--------------------------------
Experiment 1

Based on the partial support for my model found in Pilot Test 2 I conducted a larger scale online experiment that slightly modified its methodology. The most important modification is that Experiment 1 occurred in two waves, four weeks apart, in order to measure participants’ pre-existing moralization of tobacco advertising and thereby avoid concerns of reverse causality. With the exception of this preceding wave, Experiment 1 unfolded in the same fashion as Pilot Test 2 for participants. The only other differences are that payment rates on Mechanical Turk are different from those in the School’s paid participant pool, so adjustments were made to the incentives to make them more consistent with going market rates. Additionally, I employed two coders who were blind to experimental condition and the hypotheses to rate the quality of the messages that participants produced.

Participants and Procedure

Participants. In order to get a large sample I recruited participants from Amazon.com’s Mechanical Turk worker pool, which is an efficient and effective way of collecting survey and behavioural data (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). One thousand three US residents from this online labour market signed up to “join the creative marketing ideas team” in the first wave of this study. As is typical of samples drawn from this pool of workers, demographic characteristics were much more diverse than in the student samples: 49 percent of participants were female, years of age ranged between 18 and 76 ($M = 32.26, SD = 11.10$), and although the majority of participants (82.1 percent) reported a European ethnic background, 6.8 percent of participants were African American, 7.3 percent were East
Asian, 5.5 percent were Hispanic, 2.2 percent South Asian, and 0.6 percent reported other ethnic origins.

Of these 1,003 participants, 516 responded to a subsequent email with the subject line “Earn up to 5 dollars on this Creative Marketing Ideas Study” by clicking on a hyperlink embedded in the email text. Seven participants abandoned the study before viewing any pages which informed them of Philip Morris’ sponsorship of the study and a further 75 participants were excluded from analysis because they spent fewer than ten seconds reading either of two critical instruction pages. As the median time taken to read these two instruction pages was 39.5 and 84 seconds, this exclusion criteria was deemed appropriate because the manipulation was text based and liable to be ineffective if unread.\footnote{None of the results reported below were substantively altered on the basis of this exclusion criteria.} Finally, six additional participants reported technical difficulties completing the online study which interfered with their completion of the advertising task. The remaining 428 participants had almost exactly the same demographic characteristics as the original sample: 50.1 percent were female, the age range was unchanged ($M = 34.3, SD = 11.9$), and 83.4 percent of participants reported a European ethnic background, 6.4 percent an African American, 5.5 East Asian, 5.5 Hispanic, 2.1 South Asian, and 0.5 percent other.

**Procedure.** In Wave 1, when participants were first recruited, they were invited to complete a “brief 5-minute survey to register as a panelist for future studies that typically pay around $5.” During this survey participants were told that they were enrolling as a panelist in a business school’s Creative Marketing Ideas institute and that they would be anonymously contacted from time to time via the email account they had registered with Mechanical Turk
whenever a corporate-sponsored marketing study became available. In order to enroll as panelists, participants were asked to fill out a few brief questionnaires as part of their application and compensated twenty cents for their time. These questionnaires included basic demographics, the frequency with which participants use tobacco products, and the extent to which participants moralize the advertising of tobacco products. Four weeks later, once participants’ memory of these initial questionnaires had presumably decayed, participants were sent an email message notifying them that there was a new marketing study that paid up to five dollars in which they were eligible to participate. Once participants clicked the hyperlink in the text of the email they were assigned to the experimental condition of either piece rate or flat fee payment for their work and the study unfolded as had Pilot Test 2.

**Manipulation and Measures**

**Moderating variables measured in Wave 1.** After participants agreed to the consent form and had filled out demographic questions they completed the following questionnaires in a random order. The key measures regarding tobacco use and the moralization of tobacco advertising were embedded in longer questionnaires in order to disguise their importance and reduce the likelihood that these particular questions would be remembered by participants during Wave 2. At the beginning of the survey and at the top of each page participants were informed that their answers to these questionnaires would not affect their eligibility for the subsequent studies so they should answer honestly.

**Personal use of tobacco products and social media.** One questionnaire simply asked “how often do you personally engage in the following behaviors?” (1, “Never,” to 7, “At least once a day”). Embedded in a list of 12 behaviours that were chosen to appear relevant to marketing and advertising (e.g., “shop online”, “buy clothing for yourself or others,” etc.) was
the item “use tobacco products.” Overall, 19.8 percent of those who participated in Wave 2 reported using tobacco either daily or multiple times per week, which is not significantly different from the Center for Disease Control’s 18.1 percent estimate for the US population, $z = 0.910, p = .179$ (Agaku, King, & Dube, 2014). In addition, three items measured use of the social media platforms Facebook, Twitter, and an open-ended “other social media services” category, because the dependent variable was modelled on the Twitter format with which participants could be more or less familiar. Thirty-three percent of those who participated in Wave 2 reported that they never used Twitter, 32.4 percent reported using the service on a daily or weekly basis, and the remaining third reported using twitter occasionally.

**Moralization of tobacco advertising.** Another questionnaire was adapted from the Moralization of Everyday Life Scale (Lovett, Jordan, & Wiltermuth, 2012) and measured the extent to which participants moralize various advertising practices (e.g., “Advertising fast-food products designed for children” and “Corporate sponsorship of public spaces and events”). Two of these items, “Advertising tobacco products” and “Advertising products associated with negative health consequences (e.g., alcohol and tobacco products)” measured moralization of tobacco advertising on a 5-point Likert-type scale (1, “Not at all wrong (perfectly OK), to 5, “Very wrong”). The highly significant positive correlation between these two items, $r(1001) = .75, p < .001$, indicated that it was appropriate to average them to create a moralization of tobacco advertising index. Participants reported substantial variation in the extent to which they moralized tobacco advertising ($M = 2.61, SD = 1.28$), and there was a slight but significant positive skew to the data, $z_{skewness} = 2.05, p = .02$.

**Personality.** A brief 10-item measure of the Big-Five personality traits (Gosling, Rentfrow, & Swann, 2003) was added to further conceal Wave 1’s true purpose of measuring
participants’ pre-existing moralization of tobacco advertising. No specific hypotheses were made with regard to this measure and it is not discussed further.

**Wave 2 manipulation of pay structure.** Participants were randomly assigned to receive either a flat payment of $5 or pay-for-performance in the form of $1 plus $0.25 per advertising message they produced, up to a total of 16 messages. These two payment structures have equal expected value, in that participants in the pay-for-performance condition could earn a maximum of $5 if they produce 16 advertising messages. Participants were informed about their pay structure twice on two consecutive instruction pages immediately preceding the advertising task and acknowledged their awareness of the pay structure by ticking a check box.

**Dependent variables measured in Wave 2.**

**Task performance.** The same objective measure of task performance from Pilot Test 2 was used in Experiment 2, namely a count of the number of pro-tobacco advertisements written by participants. The mean number of pro-tobacco advertising messages written by participants was 5.54, the median 4, and the standard deviation 5.88. This very large range in variance is the product of a bimodal distribution pattern, whereby the most common performance was writing zero tweets (n=150), and the second most common was writing 16, the maximum number of tweets (n=75). The experimental task produced polarized performance results among participants, suggesting that in addition to treating the dependent variable as continuous it should also be examined with categorical and nonparametric methods.

**Task effort.** The same objective measure of task effort from Pilot Test 2 was used in Experiment 2, namely the amount of time spent working on the advertising task. Two extreme outliers were recorded as having spent 548.3 and 1389.2 minutes completing the task. It was assumed that these participants had left their computers during the task and returned sometime
later to complete it, so their times were Winsorized to reduce their influence on results. Participants spent a mean of 12.96 minutes completing the task, median of 7.01 minutes, with a standard deviation of 23.22 minutes, which reflects not only the large positive skew in the data but also the fact that many participants did not complete the task and therefore were coded as having spent zero minutes on it. To correct for this positive skew, the time data was log transformed as in Pilot Test 2.

**Task performance quality.** In addition to these objective measures, because advertising messages are generally considered to be more effective if they are attention-grabbing and persuasive, these qualities of the messages were coded by two trained research assistants blind to condition and hypotheses. Each message was assigned a quality score that ranged from 0, “Anti-tobacco or irrelevant or nonsensical message,” to 3, “Pro-tobacco message which exhibits high levels of creativity or persuasiveness” (see Appendix B for the coding protocol). A quadratic-weighted kappa score (Agresti, 2002) of .649 indicated that agreement between the raters was substantial according to the interpretation of kappa statistics suggested by Landis and Koch (1977). Therefore the message quality scores of the two raters were averaged and a sum of each participants’ scores was calculated to create a quality-weighted performance score that ranged between zero and 36 ($M = 10.68, SD = 11.60$). As the quality-weighted performance score is partially derived from the number of advertising messages written by participants, it has a similar bimodal distribution to the task performance variable and was therefore approached with the same analytic strategy.

**Donation opportunity.** After completing all of the questionnaires and demographic measures participants were presented with the opportunity to make an anonymous donation to the American Cancer Society. Participants were told that they could donate any percentage of
their earnings from Wave 2 of the experiment and that this donation would be automatically deducted from what they would be paid. On average, participants donated 5.67 percent of their earnings, but the large majority of participants donated nothing such that the median donation was zero and the standard deviation was 17.14, giving the distribution a strong positive skew. Furthermore the percentages donated were sporadically distributed, with clusters at ten percent, 20 percent, 50 percent and 100 percent, again suggesting that categorical and nonparametric data analysis was called for.

**Mediating variables measured in Wave 2.** The following mediating variables were measured immediately after participants completed the advertising task and the comprehension check and suspicion probe. Mediating variables were not measured before the advertising task because there was concern that they may reveal the true nature of the experiment to participants prematurely. Except for the measures of affect, which appear after the other mediating variables, the order of the following items was randomized and intermixed.

**Perceived authority.** The same three-item measure of perceived authority from Pilot Test 2 was administered again. As before, these three items exhibited high reliability, $\alpha = .80$, and thus were averaged together to create a perceived authority index.

**Perceived personal responsibility.** The same two-item measure of perceived personal responsibility from Pilot Test 2 was administered again. As before, these two items were highly and significantly correlated, $r(326) = .63$, $p < .001$, and thus were averaged together and reverse scored to create a perceived personal responsibility index.

**Perceived personal benefit.** The same two-item measure of perceived personal benefit from Pilot Test 2 was administered again. As before, these two items were highly and
significantly correlated, $r(326) = .65, p < .001$ and therefore averaged and reverse scored to create an index of perceived personal benefit.

**Affect.** The same two-item measure of guilt deployed in Pilot Test 2 was administered again in Experiment 1. As before, the positive and significant correlation between these two items, $r(326) = .75, p < .001$, indicated that it was appropriate to average them into an experienced guilt index.

**Moderating variables measured in Wave 2.** After the above listed mediating variables, the following moderating variables were presented in the order below.

**Moral attitude towards principal.** The same three-item measure of moral attitude towards Philip Morris from Pilot Test 2 was used in Experiment 1. As before, the reliability of these four items was high, $\alpha = .96$, and so they were averaged and reverse scored to create an index of participants’ moral evaluation of Philip Morris International, the purported employer in this study.

**Moralization of tobacco advertising.** The same two items used to measure participants’ moralization of tobacco advertising in Wave 1 of Experiment 1 were administered at the end of Wave 2. As in Wave 1, the correlation between these two variables was significant, $r(326) = .86, p < .001$, and so they were averaged together to create the same moralization of tobacco advertising index. Interestingly, test-retest reliability of this index was significantly lower than the internal reliability of these measures, $r(326) = .51, p < .001$, indicating that there was some degree of change in the extent to which participants moralized tobacco advertising between waves. Although this could simply indicate instability in participants’ moralization of tobacco advertising overtime, it is also possible that the experimental task of composing advertisements for a tobacco company caused some participants to change their reported moralization of
tobacco advertising in order to reduce cognitive dissonance (Festinger, 1957). Consistent with this latter interpretation, although moralization of tobacco advertising in Wave 1 predicted moralization in Wave 2, $\beta = .59$, $t(323) = 8.21$, $p < .001$, while the number of ads written by participants had no main effect on their moralization of tobacco advertising expressed in Wave 2, $\beta = .01$, $t(323) = 0.10$, $p = .918$, these factors significantly interacted, $\beta = -.22$, $t(323) = -1.92$, $p = .056$, such that those participants who wrote more ads were more likely to downgrade their previous moralization so as to make it more consistent with their behaviour in the experiment. Therefore, because the experimental task changed participants’ reported moralization of tobacco advertising, measuring their moralization several weeks before the actual experiment is a valuable contribution to the internal validity of Experiment 1.

**Moral identity centrality.** As in Pilot Study 2, Aquino and Reed’s (2002) moral identity scale was administered after participants had answered all of the other items listed above. Both the internal and symbolic moral identity subscales exhibited high reliability, $\alpha = .82$ and $\alpha = .86$, respectively.

**Comprehension check and suspicion probe.** Immediately after the advertising task participants were asked two open-ended questions that were designed to surreptitiously probe the extent to which they understood their assigned payment structure and believed the study’s cover story. These questions were “How much do you expect to be paid for completing this study?” and “Who funded this study?” Participants in the flat fee payment condition who indicated an expected payout other than $5 were considered to have misunderstood the pay structure ($n = 7$), whereas participants in the piece rate condition were given greater leeway for calculation errors due to the added complexity of determining their payout, such that only those who indicated an expected payout that was more than $2 different than their earned payout were
excluded for failing the manipulation check (n = 16; $2 was determined as an appropriate cut point on the basis of a box-and-whisker plot for identifying outliers). The answers of four participants indicated some level of disbelief in Philip Morris’s sponsorship of the study and thus these participants were removed from the analyses.

Results

Bivariate correlations between all of the aforementioned variables from Experiment 1 are displayed in Table 3. As the primary dependent variable was a count of the number advertising messages written by participants, it was first analyzed using a Poisson regression model. However, a Poisson regression model with the independent variables of tobacco advertising moralization reported in Wave 1, pay structure, and their interaction did not fit the data well, \( \chi^2(425) = 2911, p < .001 \), likely because the number of pro-tobacco advertisements written by participants was over dispersed, with a conditional variance that was greater than the conditional mean. Therefore a negative binomial regression model was used to accommodate this overdispersion and the goodness-of-fit deviance statistic improved, \( \chi^2(425) = 479, p = .033 \), although it was still significant, indicating a degree of misfit between the data and the negative binomial model. Ultimately, the results of the negative binomial model did not support my hypotheses: although participants’ moralization of tobacco advertising expressed at Wave 1 did reduce the number of advertising messages they wrote in Wave 2, \( b = -.13, SE = .06, \) Wald \( \chi^2 = 4.50, p = .034 \), there was no effect of performance-based-pay, \( b = -.11, SE = .24, \) Wald \( \chi^2 = 0.19, p = .661 \), nor a significant interaction between these terms, \( b = .012, SE = .08, \) Wald \( \chi^2 =

---

6 None of the results reported below were substantively altered on the basis of this exclusion criteria.
7 Ibid.
8 A deviance-scale standard error correction was performed, but this did not substantively change any of the reported results.
0.02, $p = .882$. Controlling for participants’ tobacco and Twitter use, as well as a host of other covariates, did not substantively alter these results.

--------------------------------

Due to the fact that the number of advertising messages written by participants was bimodally distributed and therefore did not fit the assumptions of Poisson or negative binomial regression models ideally, I also examined my data with categorical and non-parametric approaches. First I decomposed the multiple different behaviours that could be coded as a participant having produced zero advertising messages to see if any meaningful patterns were revealed. Of the 150 participants who wrote zero messages, 91 abandoned the experiment after reading that it was sponsored by Philip Morris, 31 did not write any advertising messages but completed the experiment, and 28 wrote anti-tobacco messages in opposition to the task instructions. The likelihood of engaging in one of these three insubordinate behaviours was compared against the 284 participants who wrote at least one pro-tobacco message using a multinomial logistic regression model. The results of this analysis did not reveal that moralization of tobacco advertising, pay structure, nor their interaction were significantly related to the likelihood of any of these insubordinate behaviours, and neither were covariates such as participants’ tobacco or Twitter use. A binary logistic regression comparing the likelihood of engaging in any of the insubordinate behaviours combined versus the likelihood of writing one or more pro-tobacco message was also conducted. Consistent with the negative binomial regression model, this analysis revealed that the extent to which participants moralized tobacco advertising negatively predicted their probability of writing any pro-tobacco
advertising, \( b = -0.26, SE = 0.12 \), Wald \( \chi^2 = 4.79 \), OR = 0.77, \( p = .029 \), but that neither performance-based-pay, \( b = -0.31, SE = 0.49 \), Wald \( \chi^2 = 0.40 \), OR = 0.73, \( p = .525 \), nor its interaction with moralization, \( b = -0.04, SE = 0.16 \), Wald \( \chi^2 = 0.06 \), OR = 0.96, \( p = .812 \), had any effect.

Finally, because the bimodal distribution observed in the count of pro-tobacco messages written by participants did not nicely fit the assumptions of the aforementioned models, a nonparametric test was attempted. By categorizing participants according to whether they moralized tobacco advertising or not (i.e., those participants who selected the response “1, Not at all wrong (perfectly OK)” on the moralization scale were coded as not moralizing tobacco advertising at all), I was able to create four groups of participants. Overall, 100 participants reported no moralization of tobacco advertising in Wave 1, and 52 of these were randomly assigned to receive a flat pay structure while 48 received a piece rate pay structure. Furthermore, 145 of the 335 participants who reported some moralization of tobacco advertising were assigned to the flat pay structure and the remaining 190 participants to the piece rate pay structure. Unfortunately, a Kruskal-Wallis test revealed that there was no difference in the ranked task performance of these four groups, \( \chi^2(3) = 4.35, p = .226 \), providing no evidence that pay structure interacted with moralization of tobacco advertising to affect the number of advertising messages written by participants.

In addition to counting the number of pro-tobacco advertising messages written by participants, I analyzed the summed quality scores of their advertising messages, as judged by the trained coders, as well as the amount of time participants spent on the advertising task and the amount of their earnings that they donated to the American Cancer Society. While the task performance quality score is partially derived from the task performance score, and therefore
has a similar bimodal distribution, it is not a count variable and therefore not suitable for Poisson or negative binomial regression models. I therefore submitted participants’ performance quality scores to an ordinary least squares regression, and found results that were consistent with the objective task performance data. While moralization of tobacco advertising was significantly negatively related to participants’ quality-weighted performance scores, $\beta = -0.16$, $t(425) = -2.33$, $p = .020$, neither performance-based-pay, $\beta = -0.08$, $t(425) = -0.76$, $p = .449$, nor its interaction with moralization, $\beta = 0.07$, $t(425) = 0.565$, $p = .573$, showed any effects on the quality of participants performance. None of the anticipated effects were observed in the other dependent variables either. Consistent with the quantity and quality of participants’ performance on the advertising task, the amount of time and effort they put into the task was negatively predicted by their moralization of tobacco advertising at Wave 1, $\beta = -0.15$, $t(425) = -2.152$, $p = .032$, but not their assigned pay structure, $\beta = -0.02$, $t(425) = -0.19$, $p = .842$, nor the interaction between these two factors, $\beta = -0.04$, $t(425) = -0.339$, $p = .734$. As for participants’ donation behaviour, none of the predictor variables nor their interaction were significant predictors, all $\beta$s < 1, and this was not altered by controlling for the amount that participants earned nor the number of advertising messages they wrote during the experiment.

Similar patterns of results were also observed with the intended mediator variables of Experiment 1. Ordinary least squares regression models, with tobacco advertising moralization, pay structure, and their interaction as predictors, revealed no significant main effects or interactions for perceived authority of Philip Morris in the task, perceived personal benefit, or

---

9 Time spent competing the advertising task was log transformed to correct for its significant positive skew, and the times of two outlying participants were Winsorized to prevent their undue influence on the model. The same pattern of results holds with the inclusion of the non-Winsorized outliers, but all predictors are rendered insignificant without the log transformation.
guilt, all $\beta$s < 1. The exception to these findings was participants’ sense of personal responsibility for the advertising task. The extent to which participants moralized tobacco advertising in Wave 1 negatively predicted their sense of personal responsibility during the advertising task, $b = -.224$, $t(319) = -1.980$, $p = .049$, and although performance-based-pay did not exert a main effect, $b = -.183$, $t(319) = -.411$, $p = .681$, it did significantly interact with moralization to predict participants’ perceived personal responsibility for the advertising task, $b = 0.360$, $t(319) = 2.333$, $p = .020$. I probed the nature of this interaction using a simple slopes test (Aiken & West, 1991), and found that although pay structure had no effect on participants’ sense of personal responsibility if they did not greatly moralize tobacco advertising (1 standard deviation below the sample mean), $b = 0.296$, $t(319) = 1.11$, $p = .269$, those participants who viewed tobacco advertising as more immoral (1 standard deviation above the mean moralization in the sample) felt a greater sense of personal responsibility when receiving performance-based-pay as opposed to a flat fee for their work, $b = 1.27$, $t(319) = 2.40$, $p < .001$ (see Figure 8). While this is consistent with my theorizing that performance-based-pay increases one’s sense of personal responsibility for unethical work tasks, this finding must be interpreted with caution in light of Experiment 1’s panoply of null results. Finally, it should be noted that the same pattern of null results was observed when using participants’ moralization of tobacco advertising reported in Wave 2, as well as their moral evaluation of Philip Morris, as the moderating variables in these models.
An additional moderating variable of potential importance, but about which I did not make a priori hypotheses is moral identity centrality, or the extent to which a central component of one’s identity is moral self-concept. An assumption of my model, which draws on Mazar, Amir, and Ariely’s (2008) self-concept maintenance theory, is that people care about their moral self-concept. Research by Aquino and Reed (2002), however, suggests that people vary in the extent to which they treat their moral self-concept as central to their identity. People for whom moral self-concept is a relatively unimportant component of their identity are unlikely to feel the same aversion to acting contrary to society’s hypernorms and therefore may also feel less need to morally disengage when presented with opportunities to benefit by violating those norms. According to this logic, a 3-way interaction could be hypothesized such that the effect of pay structure would only be moderated by task moralization for those people with elevated moral identities. This interaction was tested in the aforementioned negative binomial, logistic, and ordinary least squares regression models, but no evidence for it emerged. Instead, consistent with the findings of Pilot Test 2, a significant 2-way interaction between internal (but not symbolic) moral identity centrality and pay structure came out repeatedly in my results.

With regard to participants’ quantitative performance on the advertising task, a negative binomial model with pay structure, internal moral identity centrality, and their interaction term as predictors fit the data quite well, according to the deviance statistic $\chi^2(321) = 327, p = .397$.\(^{10}\)

\(^{10}\) Note that there are significantly fewer participants included in the models with moral identity as a predictor because this variable was only measured at the end of Wave 2, and which point about 100 participants had quit the survey. All of the previous analyses with moralization of tobacco advertising as a predictor were run with a filter which removed these participants from the analysis so that the sample for both analyses would be comparable, but this did not substantively change any of the reported results, suggesting that the discrepancies between the models using tobacco advertising moralization and moral identity centrality are not the result of a selection or sampling artifact.
According to this model, although internal moral identity centrality was not in itself a significant predictor, $b = 0.168, p = .10$, the significant positive effect of performance-based-pay on the number of ads written by participants, $b = 1.862, p = .026$, was moderated by the centrality of participants’ internal moral identity, $b = -.319, p = .016$. The nature of this interaction can be seen in Figure 9, and according to simple slopes tests, the performance of participants with internal moral identities that were relatively less central (1 standard deviation below the mean moral identity of the sample) was unaffected by pay structure, $b = .149, t(321) = 0.897, p = .371$, while performance-based-pay significantly reduced the number of advertising messages written by participants with internal moral identities that were relatively more central (1 standard deviation above the sample mean), $b = -0.406, t(321) = -2.46, p = .020$. Simple slopes within assigned pay structure also revealed trends such that participants with more central moral identities wrote fewer pro-tobacco advertising messages than participants with less central moral identities under pay-for-performance, $b = -0.15, t(321) = -1.80, p = .079$, but more advertisements under the flat fee structure, $b = 0.17, t(321) = 1.64, p = .102$. All of these findings are robust to the inclusion of covariates such as frequency of personal tobacco and Twitter use, as well as the inclusion or exclusion of participants flagged by attention, comprehension, and suspicion probes.

--------------------------------

Insert Figure 9 about here

--------------------------------

The same pattern of results is also observed in the quality-weighted performance scores of participants. Ordinary least squares regression revealed that both internal moral identity centrality, $\beta = .18, t(321) = 2.112, p = .035$, and pay-for-performance, $\beta = 1.26, t(321) = 3.15, p$
.002 had positive main effects on the quality of participants’ performance. Importantly, these effects were moderated by an interaction between these factors, $\beta = -1.36$, $t(321) = -3.35$, $p = .001$, as shown in Figure 10. Again simple slopes analyses revealed that pay-for-performance did not have a significant effect on the quality-weighted performance of participants with moral identities that were less central, $b = 2.72$, $t(321) = 1.51$, $p = .132$, but that it had a negative effect on the performance of participants with moral identities that were more central, $b = -5.83$, $t(321) = -3.23$, $p = .001$. Simple slopes within assigned pay structure further revealed that participants with more central moral identities significantly underperformed those with less central moral identities under performance-based pay, $b = -2.58$, $t(321) = -2.67$, $p = .008$, but that this pattern was reversed under the flat pay structure, $b = 2.55$, $t(321) = 2.28$, $p = .023$, such that participants with greater moral identity centrality actually outperformed participants with lower moral identity centrality under flat pay structures. Similarly, the amount of time that participants spent on the advertising task was positively predicted by both internal moral identity centrality, $\beta = 0.16$, $t(321) = 1.94$, $p = .053$, and to a lesser extent pay-for-performance, $\beta = .68$, $t(321) = -1.69$, $p = .093$, and the negative interaction term, $\beta = -.82$, $t(321) = -2.02$, $p = .044$ indicated that pay-for-performance had a uniquely negative impact on the effort of those participants with higher moral identities.\footnote{Time spent competing the advertising task was log transformed to correct for its significant positive skew, and the times of two outlying participants were Winsorized to prevent their undue influence on the model. The effect of internal moral identity centrality, $\beta = .16$, $t(321) = 1.95$, $p = .052$, pay-for-performance, $\beta = .65$, $t(321) = 1.63$, $p = .11$, and their interaction, $\beta = -.79$, $t(321) = -1.95$, $p = .052$, remain statistical trends in the hypothesized direction with the inclusion of the non-Winsorized outliers, but all predictors are rendered insignificant without the log transformation.} Although the above mentioned effects on quality-weighted performance scores are robust to the inclusion of personal tobacco and Twitter use covariates, the effects on time spent on the task reduced to trends in the presence of these controls. It should also be noted
that the same regression models did not significantly predict participants’ donation behaviour at the end of the experiment, with or without controls or interaction terms for the number of messages written and the payment earned by participants.

--------------------------------
Insert Figure 10 about here
--------------------------------

With respect to the proposed mediators for Experiment 1, the regression model discussed above was only able to predict participants’ sense of guilt upon completing the task, but not their perception of the employment relationship as authority ranked, their sense of personal responsibility, nor their sense of personal benefit from the task, all $\beta$s < 1. Ordinary least squares regression models predicting participants’ guilt upon completing the task revealed that internal moral identity centrality, $\beta = -0.19$, $t(321) = -2.05$, $p = .026$, and pay-for-performance, $\beta = -0.70$, $t(321) = -1.98$, $p = .084$ were significant and trending negative predictors, respectively, and that these main effects were moderated by their interaction term, $\beta = 0.76$, $t(321) = 1.85$, $p = .065$. Importantly, all of these trends became more statistically significant when controlling for the number of pro-tobacco messages written by participants, suggesting that guilt was not merely a response to the participants’ own behaviour in the experiment, but rather a product of their moral identity and the pay structure to which they were assigned. Simple slopes analyses revealed a pattern that appeared to reflect the pattern observed in participants’ performance data: the guilt of participants with less central internal moral identities was unaffected by

12 In the ordinary least squares regression model including the number of pro-tobacco messages written by participants as a control, moral identity centrality, $\beta = -0.20$, $t(320) = -2.41$, $p = .017$, performance-based pay, $\beta = -0.81$, $t(320) = -1.98$, $p = .049$, and their interaction, $\beta = 0.87$, $t(320) = -2.10$, $p = .036$, were all significant predictors of participants guilt, while the number of pro-tobacco messages written, $\beta = 0.09$, $t(320) = 1.45$, $p = .135$, was not.
performance-based-pay, \( b = -0.22, t(321) = -0.87, p = .343 \), but those participants with more central moral identities felt guiltier when paid for their performance rather than a flat fee, \( b = 0.49, t(321) = 2.06, p = .040 \) (see Figure 11). Simple slopes within pay structure also revealed that although participants with more or less central moral identities who received performance-based pay felt equally guilty, \( b = 0.07, t(321) = 0.483, p = .629 \), under the flat pay structure participants with more moral identity centrality felt significantly less guilty than did participants with moral identities that were less central, \( b = -0.41, t(321) = -2.42, p = .016 \). Nevertheless, despite this mirror-like effect, bootstrapped tests of conditional indirect effects of performance-based-pay interacting with internal moral identity through guilt did not confirm any mediation. Using model number 8 of the PROCESS macro (Hayes, 2013) with bootstrapped samples of 10,000, bias corrected 95% confidence intervals for this model’s prediction of the number of ads written by participants [-0.026, 0.463], their quality-weighted scores [-0.020, 1.106], and amount of time spent on the experimental task [-0.001, 0.065] all crossed zero, indicating that the conditional indirect effect was not significant.

Discussion

Unfortunately the results from Experiment 1 did not support my hypotheses. Of course null results concerning the interaction between pay structure and participants’ moralization of tobacco advertising are impossible to interpret definitively, but perhaps discussing the various possible explanations for this failure may be instructive. On the one hand, while erroneous theorizing, false-positive results in Pilot Test 2, and false-negative results in Experiment 1
cannot be ruled out, the fact that this relatively well powered experiment failed to replicate the supportive findings of Pilot Test 2 may point to methodological or population differences between the two studies.

The largest methodological difference between Pilot Test 2 and Experiment 1 was the multi-wave design. This allowed me to subtly measure participants’ moralization of tobacco advertising before they completed the advertising task in the second wave of Experiment 1. As predicted and discussed above, participants changed their reported moralization of tobacco advertising between the waves to be more consistent with their behaviour in the advertising task (Festinger, 1957). Thus measuring participants’ moralization of tobacco advertising several weeks before they engaged in the advertising task was critical to the internal validity of Experiment 1. However, it seems unlikely that this change in the timing of moralization’s measurement is likely to explain the discrepancy between the two studies because the measure of moralization taken after participants completed the advertising task in Experiment 2 showed the same pattern of null results.13 One other methodological discrepancy worth mentioning is the pay schedule for participants. Although in both studies participants were randomly assigned to receive either a piece-rate or flat pay structure, it was considered unwise to pay participants from Amazon Mechanical Turk’s worker population the same as undergraduate students because the market norms were different. Nevertheless, considering the wide variation in the number of Tweets written by participants in both studies, and similar bimodal distributions, it

---

13 Alternatively, it is possible that in Experiment 1 participants’ moralization of tobacco advertising measured after they completed the advertising task was less susceptible to internal consistency motivations than it was in Pilot Test 2 because in Experiment 1 participants had recorded their moralization on the same scale four weeks earlier.
does not appear that either suffered from restriction of range in the dependent variable, although any more subtle effects of the difference in pay between the studies may be difficult to detect.

The other major difference between the studies which is more likely to explain their discrepant results is the difference in populations. Demographically, the Mechanical Turk sample was significantly older ($M = 34.31$, $SD = 11.90$) than the undergraduate participant pool ($M = 20.34$, $SD = 2.59$), $t(556) = 12.92$, $p < .001$, featured more regular smokers (19.8%) than the undergraduate sample (6.8%), $z = 4.36$, $p < .001$, and was much less ethnically diverse (82.1% vs. 21.7% European ancestry, $z = 14.47$). This last demographic difference may be significant because many of the participants in the undergraduate population sampled in Pilot Test 2 are international students from non-North American cultures: notably 48.1 percent of the undergraduate sample reported East Asian ethnicity (vs. 5.5% in the online sample) and 17.9 percent South Asian ethnicity (vs. 2.1% in the online sample), although their immigration status was not recorded. Culture powerfully shapes normative expectations for how people relate to one another along several dimensions. Two cultural dimensions relevant to the present research are power distance, or the extent to which people accept hierarchical distributions of power, and individualism/collectivism, or the extent to which people expect tight-knit social ties bound by loyalty (Hofstede, 1980). Importantly, American culture is significantly lower in power distance and higher in individualism than many East Asian and South Asian cultures (Hofstede, Hofstede, & Minkov, 2010), and this may affect willingness to comply with an authority figure’s unethical instructions and the construal of individual agency in group contexts (Menon, Morris, Chiu, & Hong, 1999). Another possible difference between the undergraduate and online populations from which these samples were drawn is their level of previous experience with psychology studies. Recently scholars have begun to report substantial levels of nonnaïveté.
among samples drawn from Amazon’s Mechanical Turk (Chandler, Mueller, & Paolacci, 2014), which in some cases may undermine effects previously observed in that population (Rand et al., 2014). In the case of Experiment 1, the largest threat from nonnaïveté is that the cover story may not be believable to participants who are familiar with the use of deception in psychology experiments, but the suspicion probe, along with a large number of participants who refused to complete the task despite relatively large financial incentives, suggest that this was not a fatal flaw of the study.

Thus the explanation that the divergent results of Pilot Test 2 and Experiment 1 can be accounted for by methodological and participant population discrepancies is not sufficiently convincing, especially with regard to results that did replicate across the two studies. In both studies, the centrality of participants’ internal moral identity significantly moderated the effect of pay structure, such that performance-based pay was uniquely ineffective at motivating compliance with unethical instructions for participants with more central internal moral identities, but not those with less central moral identities. Although this interaction was not among my a priori hypotheses, its consistency suggests that it is worth theoretical interpretation with the aim of motivating future research. Such interpretation must be approached with caution, however, because internal moral identity was measured after the experimental task and therefore may be confounded with participants’ desire for internal consistency or social desirability. Despite this, the fact that Aquino and Reed's (2002) moral identity measure is worded to measure relatively stable traits rather than the temporary accessibility of moral identity, and that internal moral identity centrality was uncorrelated with all of the other variables in the Experiment other than participants’ symbolic moral identity, suggests that
interpreting this interaction as one between stable individual differences and a randomly assigned pay structure may be appropriate.

Perhaps therefore the interaction between the situation of being paid by an authority figure to perform unethical tasks and the individual characteristic of internal moral identity can be explained by social-cognitive theory (Bandura, 2001). With regard specifically to moral identity, Aquino, Freeman, Reed, Felps, and Lim (2009) argue that for people to regulate their behaviour in accordance with their moral identity that facet of their self-concept must be held consciously within their working self-concept because the self-concept is composed of many different roles and identities, only a few of which may be active at any one time (Skitka, 2003). Thus because moral identity centrality measures the tendency to have a chronically activated moral identity across situations, people with more central moral identities are more likely to behave consistently with their moral identities than are people with less central moral identities (Aquino & Reed, 2002; Sage, Kavussanu, & Duda, 2006). At first glance, however, the results from Experiment 1 seem at odds Aquino et al.’s (2009) findings that performance-based pay increases immoral behaviour in a two-party negotiation exercise among participants with more central internal moral identities, but not those with less central moral identities. The researchers explain these results from a social-cognitive perspective by arguing that relative to participants with less central moral identities, those with more central moral identities are likely to be guided by their chronically activated identities unless those identities are deactivated temporarily by the situational activation of other incompatible identities. Performance-based pay, they assert, activates self-interested facets of identity, which are incompatible with other-regarding moral identities (Grouzet et al., 2005; Schwartz, 1992), and thus people with more central moral
identities behave more selfishly when offered performance-based pay that rewards unethical behaviour.

It may be possible to reconcile the findings of Aquino et al. (2009) with those of Pilot Test 2 and Experiment 1 within social-cognitive theory by recalling that moral identity is itself a multifaceted knowledge structure composed of moral values, goals, traits, and behavioural scripts (Aquino & Reed, 2002). After all, morality is not a unidimensional construct (Graham et al., 2013), but is composed of several moral foundations, which can sometimes be at odds with one another. In Aquino et al.’s studies (2009), participants were faced with a choice between lying to promote their own self-interest and telling the truth, only one of which was consistent with participants’ moral identity. By contrast, in the present studies participants were forced to choose between two antagonistic moral behavioural scripts, obeying an authority figure and not harming others, either of which could potentially be compatible with moral identity. Thus it may be that people with more central moral identities are more likely to act in accordance with their moral identities, but that the specific manifestation of their behaviour depends on what part of their moral identities is most salient in a given situation. As discussed above, relational cues are one kind of situational factor which may influence which dimensions of moral identity are temporarily most salient (Rai & Fiske, 2011).

The cross-over interaction observed in Experiment 1 is consistent with this logic in several ways. First, in the absence of performance-based pay, which I argue erodes the perception that a relationship is authority ranked, participants with more central moral identities actually outperformed participants with less central moral identities on the advertising task. This result suggests that participants with more central moral identities may have felt a greater sense of obligation to obey the instructions under a flat pay structure, which may also explain why
they reported experiencing less guilt than participants with less central moral identities, whether or not one controls for the greater number of pro-tobacco advertisements they wrote. So, in the presence of a strong authority cue, people with more central moral identities may feel a greater sense of duty to obey than people with less central moral identities, even when the authority figure they are obeying is unethical.

Second, relative to the flat pay structure, performance-based pay significantly reduced the performance of participants with more central moral identities, despite the added incentives for their self-interest. This could be because pay-for-performance erodes authority and empowers a sense of individualism by making the presence of money in the relationship more salient (Vohs et al., 2008). As a result, participants in the performance-based pay condition may feel more powerful relative to the authority figure (or less powerless), compared to participants in the flat pay condition, and studies have shown that one’s subjective sense of power interacts with moral identity centrality, such that people with more (less) central moral identities act less (more) selfishly when they feel more powerful (DeCelles, DeRue, Margolis, & Ceramic, 2012). So, relative to flat pay structures, performance-based pay may degrade the sense of obligation one feels towards an authority figure and instead empower other dimensions of moral identity to guide behaviour, such as concerns for not harming others.

Third, under performance-based pay participants with more central moral identities were outperformed by participants with more central moral identities, contrary to their own self-interest. Under these conditions, participants with less central moral identities may be able to better seek their self-interest because they can do so without damaging their moral self-concept through moral disengagement (Bandura, 1999). People with more central moral identities have a lower overall tendency to morally disengage (Detert, Treviño, & Sweitzer, 2008), and the
degradation of the authority relationship due to performance-based pay may help prevent them from doing so. So, in the presence of a weak authority cue, people with more central moral identities may behave more consistently with non-obedience based components of their moral identity because they are less likely to morally disengage than people with less central moral identities.

One inconsistency with this perspective is that unlike Mulder and Aquino (2013), I did not find different levels of compensatory behaviour after engaging in the unethical task depending on moral identity centrality. Mulder and Aquino (2013) previously found that people with more central moral identities were more likely to engage in compensatory behaviour after committing moral transgressions, whereas people with less central moral identities were more likely to engage in consistent behaviour afterwards. Similar to their experimental design, I offered participants an opportunity to donate their proceeds from the unethical task to a well-recognized charity but I found no evidence of differences in donation as a function of the interaction between the amount of unethical work performed and participants’ internal moral identity. With the addition of the pay structure independent variable, however, this does become a search for a three-way interaction (although I also tried looking for the two-way interaction and simply controlling for the experimental manipulation), which can prove difficult to predict and reliably observe. Another relevant three-way interaction which I did not observe in my results was between the experimental manipulation of pay structure, internal moral identity centrality, and participants’ moralization of tobacco advertising. Although I would have predicted that the interaction between pay structure and moral identity would only be significant for those participants who moralized tobacco advertising this was not the case. A possible explanation for this is that participants’ moralization of tobacco advertising may be intuitively
based in moral emotions such as disgust (Rozin & Singh, 1999) and therefore poorly tapped by explicit self-report measures such as those used in this experiment, which may be confounded by more abstract values such as belief in the importance of freedom of speech and smokers’ self-determination. Future studies could construct more implicit measures of participants’ moralization of tobacco advertising, such as an implicit association test (Greenwald, McGhee, & Schwartz, 1998), in order to explore this possibility.

Obviously in light of these intriguing possibilities I regret not having measured the centrality of participants’ moral identities in Wave 1 of this experiment. However important this may appear in hindsight, at the time of design it was peripheral to my research question and I was more concerned with ensuring the internal validity of the experiment from demand characteristics and socially desirable responding. As deception was so essential to the success of Experiment 1, I intentionally limited the number of questions asking about morality and tobacco because I was concerned that participants may recall these questions and discern the true nature of the experiment. While participants’ behaviour and responses to the suspicion probe appear to indicate that I avoided one major threat to the experiment’s internal validity, this caution inadvertently compromised the internal validity of Experiment 1’s most interesting result.

**Experiment 2**

Although the results of Pilot Test 3 are difficult to interpret because of the small sample size from which they are derived, Pilot Test 3 produced no strong indications that the bug grinding paradigm is inappropriate for testing my predictions. I therefore conducted a full-scale experiment (n = 181) using the methodology trialed in Pilot Test 3 with a few minor modifications in response to the pilot’s pattern of results. The main goal of Experiment 2 was to manipulate the proposed moderator: the extent to which participants moralized the work task
they were being asked to perform. Rather than relying on pre-existing variation within the population, as was done in Experiment 1 with tobacco advertising, in Experiment 2 participants were either assigned to an experimental task that was designed to be either ethically neutral, crushing inanimate beans, or unethical, crushing living animals. This experimental manipulation of both pay structure and the ethical nature of the work task allows for a cleaner test of the proposed interaction between the two factors because random assignment eliminates the possibility of confounding third variables which may be associated with moralization. Experiment 2 also had the advantage of generalizing any findings to low-skill, repetitive work tasks, which typically benefit more from performance-based-pay than do the kind of higher-skill and creative tasks featured in Experiment 1 (Ariely, Gneezy, Loewenstein, & Mazar, 2009; Pink, 2009).

Participants and Procedure

Participants. Ninety-two community members who had previously registered in a paid participant pool administered by a large North American business school, and eighty-nine undergraduate students enrolled in introductory courses at the same school, volunteered for “A Study of Physiological Changes During Behaviour” in exchange for up to $10 in compensation, or course credit and up to $10 in compensation, respectively. This sample was 60.3 percent female, ranged in age between seventeen and fifty-nine years ($M = 21.66, SD = 5.18$), and had an ethnic composition that was 54.9 percent East Asian, 13.7 percent White, 13.7 percent South Asian, and 17.7 percent other or unspecified. Due to demographic and incentive differences between paid pool and course-credit pool populations, a dummy variable ($0 =$ paid pool; $1 =$ course-credit pool) was created to control for these differences.
**Procedure.** Experiment 2 was conducted in the same manner as Pilot Test 3, with two minor modifications (see Appendix A for the laboratory protocol provided to the male research assistant). The first modification was reducing the time allocation for participants to complete the grinding task from thirty-five to twenty-five seconds. By reducing the time participants had to complete the task, I hoped to avoid what appeared to be a ceiling effect in Pilot Test 3 in the conditions in which participants were grinding beans. Importantly, a ceiling effect in the ethically neutral conditions would reduce the ability to detect any positive effects of performance-based-pay on a simple motoric task. For the results of this experiment to be most convincing I wanted to show that this paradigm could both confirm the predictions of agency theory and find that they were reversed for unethical work tasks. The second modification I made to the procedure described in Pilot Test 3 was to employ only one male research assistant as the experimenter because authority figures are stereotypically masculine (e.g., Sczesny, 2003). Although I am not making any predictions that my theory would only apply to male authority figures, I am arguing that when it comes to laboratory experiments with small sample sizes, controlling as many extraneous variables as possible may assist in the revelation of hypothesized relationships. Participants in the laboratory only have minimal interactions with the researcher during a short time and therefore they may be more likely to draw upon stereotypes when forming impressions about the experimenter. Importantly, the theory I propose examines people’s compliance with the instructions of an authority figure, so having those stereotypes correspond to the theoretical model may produce more consistent results.

**Manipulation and Measures**

**Manipulation.**
Ethicality of experimental task. Participants were randomly assigned to place either inanimate beans (coded as 0) into a modified coffee grinder or living, captive bugs into an “extermination machine” (coded as 1), which they were lead to believe would result in the bugs’ death. Both groups of participants were told by the experimenter that it was important that they put as many objects in the grinder as they could in the 25 second time limit so that the experimenter could obtain a reliable measure of their physiological response during this time. At the end of the 25 second period participants were instructed to push the grinder’s ‘on’ button, which activated the grinder for approximately 5 seconds and provided realistic auditory feedback. The experimenter read from the same script as in Pilot Test 3 when explaining the grinding task.

Pay Structure. Participants randomly assigned to the flat-payment condition (coded as 0) were told that they would receive $10 for their participation, irrespective of their performance on the experimental task. In contrast, participants randomly assigned to the pay-for-performance condition (coded as 1) were informed that they would receive $5 irrespective of their performance, plus $0.25 for each object that they ground in the grinder. The expected utility of both conditions was exactly $10.

Manipulation check.

Moralization of the experimental task. The same manipulation check as in Pilot Test 3 confirmed that participants assigned to the unethical work task found it significantly more immoral ($M = 4.54, SD = 1.49$) than participants assigned to the ethically neutral task ($M = 2.78, SD = 1.26$), $t(180) = 8.45, p < .001$, and they also found the task significantly more immoral than the neutral scale midpoint, $t(102) = 3.71, p = .001$, indicating that the manipulation was successful.
Dependent variables.

**Task Performance.** The number of objects participants placed into the coffee grinder within the 25 second period was the primary dependent variable for this study. Unlike in Experiment 1, participants’ performance on this task was much more normally distributed. Participants placed an average of 12.69 objects in the grinder, with a standard deviation of 4.78 objects. Moreover, although there appeared to be a mean shift in performance depending on whether participants were asked to grind bugs ($M = 10.3$, $SD = 4.75$) or beans ($M = 15.78$, $SD = 2.58$), neither condition resulted in significant skew, $z_{skewness} = 0.64$, $p = .522$ and $z_{skewness} = -1.49$, $p = .136$, respectively. In light of these approximately normal distributions, it was decided that ordinary least squares regression models would constitute appropriate analysis, although negative binomial models were also conducted because task performance was technically measured as a count variable.

**Voiced protest.** The number of times participants refused to participate in the experimental task was measured as a secondary dependent variable. The maximum possible number of refusals was three, after which the experimenter was instructed to allow the participant to complete the questionnaires or withdraw from the study completely. Six participants voiced a single protest about the work task but complied when prompted by the experimenter, and a further five participants refused to complete the work task despite this scripted prompting.

**Mediating Variables.**

**Perceived authority.** The three items from Experiment 1 and Pilot Tests 2 and 3 were used to measure participants’ perception of their relationship with the experimenter as authority
ranked. These three items exhibited acceptable reliability, $\alpha = .60$, and so were averaged together to create a perceived authority index.

**Perceived personal responsibility.** The same two items from Experiment 1 and Pilot Tests 2 and 3 were used to measure the extent to which participants perceived themselves or the researchers to be responsible for the consequences of their behaviour in the experiment. These two items were positively and significantly correlated, $r(180) = .38$, $p < .001$, and averaged together and reverse scored to create a perceived personal responsibility index.

**Perceived personal benefit.** The same two items from Experiment 1 and Pilot Tests 2 and 3 were used to measure the alternative mechanism of perceived personal benefit. These items were positively and significantly correlated, $r(180) = .40$, $p < .001$ and therefore averaged and reverse scored to create an index of perceived personal benefit.

**Affect.** As in Pilot Test 3, a modified version of the PANAS (Watson et al., 1988) was employed in this study to measure guilt and disgust. Disgust was added to this study because it was thought that some participants’ disgust response to bugs may affect their performance on the experimental task. Three items intended to measure guilt, ‘remorse,’ ‘regret,’ and ‘guilt’ displayed high internal reliability, $\alpha = .89$, and were thus averaged together to create a self-reported guilt index. The two items intended to measure disgust, ‘grossed out’ and ‘disgust’ were also positively and significantly correlated, $r(180) = .79$, $p < .001$, and thus averaged into a self-reported disgust index.

**Moderating variable: moral identity.** Given that participants’ internal moral identity interacted significantly with performance-based-pay in Pilot Test 2, Aquino and Reed’s (2002) moral identity scale was added to Experiment 2 and administered after participants had
answered all of the other items listed above. Both the internal and symbolic moral identity subscales exhibited high reliability, $\alpha = .82$ and $\alpha = .79$, respectively.

**Suspicion probe.** At the end of the questionnaires participants in the unethical experimental condition were informed that they had not in reality killed any bugs using the purported extermination machine. At this point participants were asked to indicate whether they had previously known that the bugs were not really killed during the experiment. Of participants in the unethical task condition, 19.2 percent indicated that they had not believed that the bugs were actually killed in the experiment and were thus removed from analyses.\textsuperscript{14}

**Results**

Bivariate correlations between all of the aforementioned variables from Experiment 2 are displayed in Table 4. I start my analyses with an exploration of the probability that participants would voice protest about the work task they were being asked to complete. In total, 11 participants (6 percent) voiced some unwillingness to comply with the experimenter’s instructions, and five steadfastly refused despite three prompts. Because all 11 participants voicing insubordination were assigned the unethical task, fitting logistic models to explore the hypothesized interaction between pay structure and task ethicality in this data proved impossible due to singularities in the Hessian matrix. Therefore I only examined the effect of pay structure within participants assigned to the unethical work task. Multinomial logistic regression revealed that relative to a flat pay structure, pay-for-performance did not significantly increase the likelihood that participants would either voice a single complaint prior to complying, $b = .63$, $SE = .94$, Wald $\chi^2 = 0.45$, OR = 1.47, $p = .502$, nor steadfastly refuse to comply, $b = .41$, $SE = \ldots$

\textsuperscript{14} None of the results reported below were substantively altered on the basis of this exclusion criteria.
1.23, Wald $\chi^2 = 0.11$, OR = 1.33, $p = .741$. Moreover, when these two behaviours were collapsed into a single category of participants who voiced reticence regarding the work task, assignment to performance-based-pay was no more likely to lead to voicing concern than a flat pay structure, $b = -0.03$, SE = .80, Wald $\chi^2 = 0.001$, OR = 0.97, $p = .969$, according to a binomial regression. Additionally, interacting participants’ assigned payment structure with their self-reported internal moral identity did not reveal any significant predictors of voice. Pay-for-performance, $b = .96$, SE = 4.16, Wald $\chi^2 = 0.054$, OR = 2.62, $p = .817$, internal moral identity, $b = .03$, SE = .46, Wald $\chi^2 = 0.004$, OR = 1.03, $p = .950$, and their interaction term, $b = -0.23$, SE = .69, Wald $\chi^2 = 0.11$, OR = 0.794, $p = .738$, all failed to predict whether participants would speak up against the unethical instructions.

---------------------------------------------

Insert Table 4 about here

---------------------------------------------

With regard to participants’ compliance with the experimenter’s instructions, similar null results were found. As mentioned above, the distribution of participants’ task performance was approximately normal and thus only results from ordinary least squares regression models are reported, although negative binomial regression models also produced similar results. It was found that although participants ground fewer bugs than beans, $\beta = -.65$, $t(174) = -7.02$, $p < .001$, there was no main effect of pay-for-performance relative to the flat pay structure, $\beta = -.07$, $t(174) = -0.74$, $p = .458$, and there was no significant interaction between these predictors that affected performance, $\beta = .15$, $t(174) = 1.30$, $p = .195$. Importantly, taking participants’ internal (or symbolic) moral identities into account did not alter this pattern, as all standardized regression coefficients in a model including a three-way interaction between task ethicality, pay
structure, and moral identity were less than one and statistically insignificant. See Figure 12 for a graphical representation of the test for this three-way interaction. For comparability of results with Experiment 1, and because the anticipated positive effect of performance-based pay did not emerge in the ethically neutral condition, the two-way interaction between performance-based pay and internal moral identity centrality was examined in the unethical condition only.

Unfortunately, although all of the regression coefficients were in the same direction as found in the other experiments, performance-based pay, $\beta = 0.74$, $t(80) = 0.91$, $p = .365$, internal moral identity centrality, $\beta = -0.04$, $t(80) = -0.29$, $p = .773$, and their interaction, $\beta = -0.66$, $t(80) = 0.81$, $p = .422$, none of them reached significance.

Finally, participants’ responses to the predicted mediators in Experiment 2 were examined in an attempt to glean an understanding of their performance behaviour. Unfortunately, ordinary least squares regression models predicting the extent to which the relationship with the experimenter was perceived as authority ranked, participants sense of personal responsibility for and personal benefit from the outcome of the work task, as well as their feelings of guilt and disgust did not reveal any statistically significant interactions between pay structure, task ethicality, and internal or external moral identity. Only main effects were significant in these models, and although for the most part they were consistent with my theorizing, as can be seen in the bivariate correlations presented in Table 4, most of these main effects were no longer significant when interaction terms were added.

**Discussion**
While the null results from Experiment 2 are difficult to interpret, some patterns in the data suggest methodological challenges which would need to be addressed in future adaptations of this paradigm. First, with regard to the ethically neutral condition, it is surprising that pay-for-performance did not result in improved task performance, especially since tasks which are simple, repetitive, and motoric are the most sensitive to pay-for-performance enhancements (Guzzo et al., 1985; Jenkins et al., 1998; Locke et al., 1980; Pink, 2009; Stajkovic & Luthans, 1998). Although I adjusted the amount of time participants had to complete the task such that participants did not perform at the ceiling of the performance measure itself, participants may nevertheless have been performing at the ceiling of their abilities, above which any performance improvements caused by increased motivation to obtain incentives may not have been detectable. I propose this explanation because the mere presence of the experimenter observing their behaviour may have caused high levels of performance by inducing an optimal level of arousal for this simple, repetitive, and motoric task (Zajonc, 1965). In addition to this effect, it may be that the performance-based pay induced excessive arousal that was detrimental to performance, although this had not been anticipated for such a simple task (Ariely, Gneezy, et al., 2009). Although lowering the amount of performance-based pay may solve this latter problem, if it is the case that social facilitation from the experimenter’s observation of participants is causing ceiling effects, then this paradigm may not be appropriate for testing the ethical boundary of agency theory because it is unlikely to show the positive effect of performance-based pay on ethically neutral tasks.

Second, with regard to the failure to conceptually replicate the interaction between pay structure and internal moral identity centrality observed in Pilot Test 2 and Experiment 1, I think it is informative to examine participants’ performance distributions across the two experimental
paradigms. While participants’ performance in the online advertising task was bimodally
distributed in Pretest 2 and Experiment 1, with a large proportion of participants refusing to
comply with the experimental task, participants’ performance in the laboratory bug grinding
task approximated a normal distribution. This more homogenous response to the bug grinding
task suggests that participants may have felt greater pressure to comply with the authority
figure’s instructions and less freedom to refuse to obey. Indeed, Milgram (1974) found that
obedience to an authority figure is significantly enhanced by their physical proximity, which
was certainly much greater in the laboratory study compared to the online study. Importantly,
personality traits exert a smaller influence over behaviour in “strong” situations and a greater
influence in “weak” situations (Beaty, Cleveland, & Murphy, 2001; Mischel, 1977). Therefore
the authority figure’s strong influence over participants’ behaviour in the laboratory may have
overpowered the interactive effects of moral identity centrality that were observed in the two
online studies, where the authority figure necessarily more remote. Although this interpretation
suggests a boundary condition for any interactive effect between performance-based pay and
moral identity, it does not necessarily imply that such an effect would be unimportant. After all,
in many organizational settings it is impractical to have direct supervision of employees at all
times, and as technologically-mediated remote employment continues to expand across
platforms such as Mechanical Turk and oDesk, employees may often perceive the latitude to
resist the constraints of an authority relationship if they feel it is unethical.

General Discussion

Although I am disappointed that my a priori hypotheses were not confirmed by
Experiments 1 and 2, I am heartened by the fact that failure is part of the scientific process.
Under the guidance of experts in experimental methodology I took careful steps to construct
valid tests for my theory. These steps included conducting three pilot tests of distinct experimental paradigms, one of which was created specifically for this research program, which allowed me to identify the most appropriate paradigms and refine them in response to pilot test data. The two most appropriate paradigms were then used to conduct well-powered experimental tests across two very different work tasks, which is important because some evidence suggests that pay-for-performance is not equally effective across all tasks (Ariely, Gneezy, et al., 2009; Pink, 2009). Furthermore, both experiments included comprehension, manipulation, and suspicion checks to assess internal validity and guard against demand characteristics and socially desirable responding. These checks provide confidence that the deception involved in these experiments produced the genuine behaviour I planned to study. Although null-results cannot be interpreted as disconfirming my theory, hopefully my efforts have shown that producing evidence supporting an interaction between performance-based pay and the subjective ethicality of a work task is not easily accomplished.

Another important part of the scientific process is the discovery of novel and unexpected phenomena. I began this research inspired by Milgram's (1974) experiments, which appear to show how situational forces can drive the unethical behaviour of ordinary people to levels of extraordinary darkness. In my search for organizationally relevant factors which might shed greater light on destructive obedience and how it may be alternately propagated or prevented, I assumed, perhaps implicitly, that situations determine moral behaviour (Doris, 1998; Harman, 2003), and neglected to more seriously consider stable personality traits which were of potential theoretical relevance. In light of my findings in Pilot Test 2 and Experiment 1, however, I am now much more cognizant of the explanatory power of the interaction between individual differences and situations when it comes to understanding (un)ethical behaviour. I therefore
concur with calls for greater research into this relatively unexplored domain of behavioural ethics research (Aquino et al., 2009; Hardy & Carlo, 2005; Walker & Frimer, 2007), and propose that as the first study examining the relationship between moral identity and obedience to an unethical authority figure this research makes a small but potentially meaningful contribution to this literature.

My research suggests that the effect of moral identity on behaviour depends on the relationship within which that behaviour is embedded. Surprisingly, because obedience to an authority figure can be a moral obligation (Graham et al., 2011, 2013), people with more central moral identities may be more likely to obey than people with less central moral identities, even when an authority figure issues instructions that may result in the harming of others. This suggests that moral identity may itself be multidimensional, and that although people with more central moral identities are more likely to be guided by ethical principles, which particular principles they end up following may be situationally determined. Indeed, the importance of the situation for determining the particular content of moral identity which is activated and influences behaviour is suggested by the effect of performance-based pay on compliance with unethical instructions. Uniquely among those with more central moral identities, but not those with less central moral identities, performance-based pay had the surprising effect of reducing compliance with an unethical request, presumably because such incentives undermine normative authority (Fehr & Falk, 2002; Fehr & List, 2004; Grant, 2012) and enhance one's sense of individual independence (Vohs et al., 2008). Whereas individuals with less central moral identities may be motivated to pursue their own self-interest in the light of performance-based incentives, and displace responsibility onto the person issuing unethical instructions as a means of maintaining their moral self-regard, those with more central moral identities appear less likely
to morally disengage in the presence of such a degraded authority relationship and instead follow ethical principles such as the proscription against harming others.

Of course these conclusions are limited by the fact that they have been derived post hoc from a pattern of unanticipated results. Furthermore, the measurement of moral identity centrality after the completion of the unethical task leaves these conclusions open to questions about their directionality and alternate explanations, such as the desire for consistency between thoughts and behaviours. Nevertheless, this research produced some evidence that this relationship is not entirely spurious and may warrant further investigation. The fact that this interaction replicated, albeit with somewhat different patterns, across Pilot Test 2 and Experiment 1, suggests that this effect is reliable and generalizable across fairly different populations. In addition, the fact that internal moral identity centrality did not have significant bivariate correlations with either the independent or dependent variables, and that the result of interest is an interaction, make parsimonious alternative explanations less likely. For example, participants’ desire for internal consistency (Festinger, 1957) cannot easily explain why in Experiment 1, depending on the pay structure to which they were assigned, participants reporting more central moral identities wrote both significantly more and significantly fewer pro-tobacco advertising messages than participants reporting less central moral identities.

In light of these promising results, researchers may want to continue exploring the interaction between situational factors and personality as it relates to obeying unethical instructions. Especially in light of the theoretical arguments posed here, it would seem pertinent to examine the interactive effects of stable individual differences such as the extent to which people endorse the five intuitive foundations of morality (Graham et al., 2011) and the dispositional tendency to morally disengage (Moore et al., 2012), in addition to moral identity.
centrality. Other potentially interesting variables which may influence destructive obedience include duty orientation (Hannah, Jennings, Bluhm, Peng, & Schaubroeck, 2014) and power distance (Hofstede, 1980). Clearly there is still much to learn by applying a social-cognitive perspective to the study of moral behaviour, and in particular this research highlights the importance of situational cues for understanding the relationship between moral identity and obedience to unethical authority.
References


doi:10.1016/j.riob.2013.08.001


doi:10.1080/19416520.2011.574506


doi:10.1016/j.obhdp.2013.03.005


doi:10.1037/a0032740


Figure 1. Proposed theoretical model.

- Pay Structure
- Employee’s perception of employment relationship as authority ranked
- Employee’s sense of personal responsibility
- Employee’s ethical evaluation of instruction
- Compliance Behavior

H1
H2
H3
H4
Figure 2. Graph of the extent to which participants lied in Pilot Test 1.
Table 1. Correlation matrix of variables from Pretest 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pay-for-performance</td>
<td>0.48</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Number of pro-tobacco advertising messages written</td>
<td>5.45</td>
<td>4.23</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Time spent on advertising task (minutes)</td>
<td>8.85</td>
<td>8.75</td>
<td>.01</td>
<td>.55***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Percent of compensation donated</td>
<td>6.36</td>
<td>4.80</td>
<td>.25*</td>
<td>-.38***</td>
<td>-.30**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Perceived authority of principal</td>
<td>4.60</td>
<td>1.66</td>
<td>-.07</td>
<td>.25*</td>
<td>.20*</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perceived personal responsibility for advertising task</td>
<td>4.12</td>
<td>1.87</td>
<td>.14</td>
<td>-.25*</td>
<td>-.25*</td>
<td>.08</td>
<td>-62***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Perceived personal benefit from advertising task</td>
<td>3.05</td>
<td>1.82</td>
<td>.06</td>
<td>-.28**</td>
<td>-.14</td>
<td>.10</td>
<td>-.89***</td>
<td>.46***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Guilt</td>
<td>3.83</td>
<td>1.77</td>
<td>.02</td>
<td>.15</td>
<td>.12</td>
<td>-.06</td>
<td>.54***</td>
<td>-.31**</td>
<td>-.52***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Moralization of principal</td>
<td>4.90</td>
<td>1.54</td>
<td>.13</td>
<td>-.24*</td>
<td>-.09</td>
<td>.23*</td>
<td>-.11</td>
<td>.09</td>
<td>.14</td>
<td>.27**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Moralization of tobacco advertising</td>
<td>5.16</td>
<td>1.37</td>
<td>.10</td>
<td>-.19</td>
<td>-.13</td>
<td>.11</td>
<td>-.06</td>
<td>.04</td>
<td>.07</td>
<td>.30**</td>
<td>.76***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Frequency of personal tobacco use</td>
<td>1.74</td>
<td>1.61</td>
<td>.02</td>
<td>.29**</td>
<td>.17†</td>
<td>-.15</td>
<td>.07</td>
<td>-.10</td>
<td>-.10</td>
<td>-.14</td>
<td>-.31**</td>
<td>-.27**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Internal moral identity</td>
<td>5.92</td>
<td>.97</td>
<td>.06</td>
<td>-.12</td>
<td>.06</td>
<td>.18†</td>
<td>.07</td>
<td>.05</td>
<td>-.10</td>
<td>.30**</td>
<td>.54***</td>
<td>.44***</td>
<td>.35***</td>
<td></td>
</tr>
<tr>
<td>13. Symbolic moral identity</td>
<td>4.51</td>
<td>1.05</td>
<td>-.17†</td>
<td>.00</td>
<td>-.09</td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
<td>-.05</td>
<td>.11</td>
<td>.15</td>
<td>.35***</td>
<td>.29**</td>
<td>.39***</td>
</tr>
</tbody>
</table>

Note. M = mean. SD = standard deviation.  
†p < .10. *p < .05. **p < .01. ***p < .001.
**Figure 3.** Objective performance on advertising task in Pilot Test 2 by pay structure and moralization of tobacco advertising.

![Graph showing objective performance on advertising task in Pilot Test 2 by pay structure and moralization of tobacco advertising.](image)

**Figure 4.** Objective performance on advertising task in Pilot Test 2 by pay structure and internal moral identity centrality.

![Graph showing objective performance on advertising task in Pilot Test 2 by pay structure and internal moral identity centrality.](image)
**Figure 5.** Perception of employment relationship as authority ranked in Pilot Test 2 by pay structure and moralization of tobacco advertising.
Figure 6. Laboratory set-up in Pilot Test 3 and Experiment 2.
Table 2. Correlation matrix of variables from Pilot Test 3.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Unethical task</td>
<td>.41</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Pay-for-performance</td>
<td>.61</td>
<td>.49</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Task performance</td>
<td>15.97</td>
<td>4.72</td>
<td>-.55***</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Moralization of work task</td>
<td>4.49</td>
<td>1.85</td>
<td>.74***</td>
<td>.25</td>
<td>-.29†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Perceived authority of principal</td>
<td>3.60</td>
<td>1.40</td>
<td>-.57***</td>
<td>-.12</td>
<td>.39*</td>
<td>-.32*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perceived personal responsibility</td>
<td>4.77</td>
<td>1.35</td>
<td>.31*</td>
<td>.27†</td>
<td>-.07</td>
<td>.30†</td>
<td>-.26†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Perceived personal benefit</td>
<td>3.88</td>
<td>1.46</td>
<td>-.01</td>
<td>.45**</td>
<td>-.02</td>
<td>-.05</td>
<td>-.41**</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Guilt</td>
<td>2.87</td>
<td>1.16</td>
<td>.48**</td>
<td>.19</td>
<td>-.18</td>
<td>.44**</td>
<td>-.10</td>
<td>.20</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>9. Disgust</td>
<td>2.82</td>
<td>1.13</td>
<td>.65***</td>
<td>.18</td>
<td>-.56**</td>
<td>.46**</td>
<td>-.32*</td>
<td>.27†</td>
<td>-.04</td>
<td>.68***</td>
</tr>
</tbody>
</table>

Note. M = mean. SD = standard deviation. †p < .10. *p < .05. **p < .01. ***p < .001.
Figure 7. Performance on grinding task in Pilot Test 3 by pay structure, ethicality of task, and gender of research assistant.
Table 3. Correlation matrix of variables from Experiment 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pay-for-performance</td>
<td>0.54</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Number of pro-tobacco advertising messages</td>
<td>5.54</td>
<td>5.88</td>
<td>-0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Logged time spent on advertising task</td>
<td>0.67</td>
<td>0.60</td>
<td>-0.06</td>
<td>0.81***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Quality-weighted performance score</td>
<td>10.81</td>
<td>11.69</td>
<td>-0.04</td>
<td>0.98***</td>
<td>0.80***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Percent of compensation donated</td>
<td>5.69</td>
<td>17.14</td>
<td>0.10†</td>
<td>-0.20***</td>
<td>-0.20***</td>
<td>-0.20***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perceived authority of principal</td>
<td>4.02</td>
<td>1.74</td>
<td>-0.25***</td>
<td>0.34***</td>
<td>0.40***</td>
<td>0.36***</td>
<td>-0.18**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Perceived personal responsibility for advertising task</td>
<td>4.14</td>
<td>1.80</td>
<td>0.20**</td>
<td>-0.23***</td>
<td>0.23***</td>
<td>0.25***</td>
<td>0.20**</td>
<td>-0.57***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Perceived personal benefit from advertising task</td>
<td>3.42</td>
<td>1.86</td>
<td>0.20**</td>
<td>-0.42***</td>
<td>0.49***</td>
<td>0.46***</td>
<td>0.20**</td>
<td>-0.75***</td>
<td>0.55**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Guilt</td>
<td>3.49</td>
<td>1.74</td>
<td>0.05</td>
<td>0.05</td>
<td>0.15**</td>
<td>0.08</td>
<td>0.02</td>
<td>0.27***</td>
<td>-0.14*</td>
<td>-0.24***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Moralization of tobacco advertising at Wave 1</td>
<td>2.66</td>
<td>1.31</td>
<td>-0.14**</td>
<td>-0.17***</td>
<td>-0.14**</td>
<td>0.15**</td>
<td>-0.09</td>
<td>0.00</td>
<td>0.13*</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Moralization of tobacco advertising at Wave 2</td>
<td>2.54</td>
<td>1.44</td>
<td>0.11*</td>
<td>-0.24***</td>
<td>-0.33***</td>
<td>-0.24***</td>
<td>0.20**</td>
<td>-0.21***</td>
<td>0.13*</td>
<td>0.29***</td>
<td>0.25**</td>
<td>0.51***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Moralization of principal</td>
<td>4.83</td>
<td>1.80</td>
<td>0.18**</td>
<td>-0.21***</td>
<td>-0.24***</td>
<td>-0.19**</td>
<td>0.13*</td>
<td>-0.21***</td>
<td>0.13*</td>
<td>0.25***</td>
<td>0.42**</td>
<td>0.32***</td>
<td>0.61**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Internal moral identity</td>
<td>6.24</td>
<td>0.87</td>
<td>0.00</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.07</td>
<td>0.10†</td>
<td>0.08</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Symbolic moral identity</td>
<td>4.43</td>
<td>1.29</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-0.10</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.19**</td>
<td>0.38**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>15. Frequency of personal tobacco use</td>
<td>2.21</td>
<td>1.94</td>
<td>-0.05</td>
<td>0.10*</td>
<td>0.11*</td>
<td>0.10*</td>
<td>-0.07</td>
<td>0.07</td>
<td>0.00</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.24***</td>
<td>-0.23***</td>
<td>-0.26***</td>
<td>0.08</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>16. Frequency of personal Twitter use</td>
<td>3.16</td>
<td>1.96</td>
<td>-0.07</td>
<td>0.11*</td>
<td>0.07</td>
<td>0.10*</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
<td>-0.09</td>
<td>0.10†</td>
<td>-0.11*</td>
<td>0.03</td>
<td>-0.03</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

Note. M = mean. SD = standard deviation. †p < .10. *p < .05. **p < .01. ***p < .001.
Figure 8. Sense of personal responsibility for the advertising task in Experiment 1 by pay structure and moralization of tobacco advertising.
Figure 9. Number of advertising messages written by participants in Experiment 1 by pay structure and internal moral identity.
**Figure 10.** Participants’ quality-weighted performance scores on the advertising task in Experiment 1 by pay structure and internal moral identity.
Figure 11. Participants’ sense of guilt after completing the advertising task in Experiment 1 by pay structure and internal moral identity.
Table 4. Correlation matrix of variables from Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Unethical task</td>
<td>0.56</td>
<td>0.50</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>2. Pay-for-performance</td>
<td>0.52</td>
<td>0.50</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>3. Voiced dissent</td>
<td>0.06</td>
<td>0.24</td>
<td>.22**</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>4. Task performance</td>
<td>12.69</td>
<td>4.78</td>
<td>-.56**</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>6. Perceived authority of principal</td>
<td>3.96</td>
<td>1.28</td>
<td>-.21**</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
<td>-.16*</td>
</tr>
<tr>
<td>7. Perceived personal responsibility</td>
<td>4.55</td>
<td>1.29</td>
<td>.13†</td>
<td>.19**</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>8. Perceived personal benefit</td>
<td>3.60</td>
<td>1.38</td>
<td>-.07</td>
<td>.35**</td>
<td>.13†</td>
<td>-.08</td>
<td>-.08</td>
<td>-.08</td>
<td>-.08</td>
<td>-.08</td>
<td>-.08</td>
</tr>
<tr>
<td>9. Guilt</td>
<td>2.79</td>
<td>1.00</td>
<td>.42**</td>
<td>.10</td>
<td>-.02</td>
<td>-.16*</td>
<td>.42**</td>
<td>.04</td>
<td>.10</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>10. Disgust</td>
<td>2.93</td>
<td>1.15</td>
<td>.51**</td>
<td>.05</td>
<td>.04</td>
<td>-.33**</td>
<td>.43**</td>
<td>-.06</td>
<td>.06</td>
<td>-.15*</td>
<td>.66**</td>
</tr>
</tbody>
</table>

Note. M = mean. SD = standard deviation.  
†p < .10. *p < .05. **p < .01. ***p < .001.
Figure 12. Performance on grinding task in Experiment 2 by pay structure, ethicality of task, and participants’ internal moral identity.
Appendix A. Laboratory protocol for Pilot Test 3 and Experiment 2.

Experimental Protocol for Fall 2014 / Spring 2015 Bug Study

1. Booking the lab and posting study sessions

1.1. Noon, 21 days before you conduct a session

1.1.1. Book lab LL1015J on R-world. When you post your booking, write “A Study of Physiological Changes During Behaviour” in the description and set the time for 30 minutes before and after the SONA spots that you plan to post. Lab space can be in high demand, so you should try to book space as early as possible. Note that just because you have booked lab space doesn’t mean that you have to schedule SONA slots at the same time – in fact, it’s better if you don’t post SONA slots more than a week in advance, because otherwise participants can forget their appointments.

1.1.1.1. Five days before you conduct the study, book the same hours in LL1015G. If this room is taken, you may also use LL1015D and just note this in the computer survey once participants have finished with the study.

1.1.1.2. We are only allowed to book 14 hours worth of lab space per week. See the ‘Lab booking policy.doc” in the main RA folder for booking last minute lab space if we need to book more than 14 hours per week.

1.1.1.3. If you are running a separate study for Chen-Bo, that study should be given a different description in SONA and then will not count towards the same 14 hour per week limit.

1.1.1.4. I think only 1 RA needs to be present to run this study. Let me know if this does not work.

1.2. Post “A Study of Physiological Changes During Behaviour” sessions on SONA’s paid pool with space for 1 participant per session.

1.2.1. These slots should be posted 7 days in advance of when the study will actually run (you may post as few as 72 hours before, but try to avoid this, and do not post with less time than this or more than a week in advance). Specify whomever is going to be conducting the study as the researcher and book until we have collected 200 participants.

1.2.2. Due to the much slower nature of the paid pool during the summer, I expect that the booking procedure will work much more like being on call. The idea would be to post as many slots during the week as you expect you could make and then only show up to those sessions for which participants sign up. I have configured the study so that the SONA system will email the specified researcher whenever someone signs up for or cancels a study appointment and I have prevented participants from signing up or cancelling an appointment 24 hours before the study is scheduled to begin.

1.2.3. Schedule sessions for every 40 minutes during time for which you are available.
1.3. The day before your session, send participants the following reminder email by logging into SONA and viewing the study information. Click on “Contact Participants,” specify participants that are signed up for sessions tomorrow, and then copy and paste this text over the text box, deleting the default message in the text box. Send a copy of the message to yourself so that you know the system sent it.

1.3.1. Dear Participant,

This is a reminder that you are scheduled for a study session in the behavioural labs tomorrow. The labs are located on the lower level of the new Rotman building (95 St. George Street), across the street from Robarts Library. Please arrive in the waiting area outside the labs at the time of your appointment, which can be found in the previous email from the SONA system or in your SONA account: https://rotman.sona-systems.com/default.aspx

Thank you.

2. 30 minutes before participants arrive, make sure the following materials are prepared.

2.1. Sign out a key for the behavioural labs at the front desk of the old Rotman building (105 St. George street). Always shut the door behind you as you enter the lab, to keep participants from wandering in early.

2.2. Get the key for the cash box and the biopac filing cabinet and retrieve the cash box. Perform a count of the cash in the box. Is this enough for your study sessions today (plan for $10 per participant). If not, call Julian immediately (919)561-1222 so that he can replenish. Log your starting count in the ‘Cash box accounting’ spreadsheet in the Bug study Folder. Leave the cash box in the cabinet with the cabinet closed.

2.3. Check what conditions you will be running today on the ‘Bug study randomization’ spreadsheet. Print out a copy of this sheet for the day and keep it in your clipboard. Each time you start a new condition tick it off on the sheet and then turn your protocol to the page with the text instructions for that condition that you will read to participants, which will help you remember which condition you are running and help things run more smoothly with each participant.

2.4. Make sure there are enough copies of the debriefing sheets on top of the cabinet to give to participants once they have completed the study.

2.5. Set the lock for lab LL1015J and LL1015G to ‘open,’ so that the door does not automatically lock every time it shuts. You will have to remember to lock the doors when you leave for the end of the day.

2.6. Configure LL1015J according to the diagram and instructions.
2.6.1. Get the laptop and charger from the top shelf of the biopac cabinet. Set it up according to the room configuration diagram. Connect the biopac receivers to it using the blue Ethernet cable. Plug into the power outlets in the wall beside the chair. Turn on the laptop but keep the biopac receiver turned off with its cables facing the wall. Load the ‘simulated monitor animation’ and pause it and minimize the window. Leave the laptop open at an easy viewing angle for you seated is the chair.

2.6.2. Place grinder slightly to the left of the power outlets in the center of the table and 6 inches back from the edge of the table with the controls facing towards where the participant will stand and plugged into the power outlet on the table. Grinder is kept under the kitchen sink.

2.6.2.1. Whenever possible, keep the funnel and pipe attached to the grinder as repeated disassembly may cause the joints to weaken and the paint to flake. Treat the pipe and funnel with care as breakage of this custom made machine will be hard to repair.

2.6.2.2. To open the grinding chamber, hold the lid in one hand and the machine body in the other, twist the lid counter-clock-wise and then lift lid, pipe and funnel off grinder as a single unit. Grinding chamber may be detached from the grinding body by twisting it counter-clock-wise. Empty used grinding medium into trash in another room where participants will not see it. Fill the grinding chamber with 10 beans and \( \frac{1}{2} \) a sheet of paper torn into about 10 pieces. Not sure how many times this mixture will sound convincing before being turned to powder, so listen each time you grind and determine whether you think this mixture needs to be replaced for the next session. At the end of the day, you should probably recharge the grinding chamber with fresh beans and eraser so that it is ready for the next day.
2.7. Get the wrist monitor out and put it in your lab coat pocket, being careful not to tangle the wires or bend them so they break. Charge the monitor during down time, such as a lunch break, but don’t leave it charging overnight. Get out a bag of EL507 sensors and a role of medical tape and place them in the control room; take out a strip and place it in your pocket. Put a roll of the medical tape in your lab coat pocket. Place the stop watch in your lab coat pocket. When not in use, keep the envelope of EL507 sealed, otherwise the gel on the stickers will dry out.

2.8. Bugs are kept on the top shelf in the control room. Even if you don’t have any bug conditions that you are running that day, take down all the containers and remove their lids to vent for about 15 minutes. Fun fact: these bugs are actually crustaceans! Therefore they need the air humidity to breath. Thus we want to keep the soil in their containers moist, but not soggy. If you notice the soil is dry, mist the soil a couple of times until it is moist. If the lid and sides of the container have condensation on them when you start the experiment for the day, the containers are too humid and there is a risk that mold/mildew will grow. To reduce the humidity, leave the lid off the container for an hour or so, allowing the excess humidity to evaporate. All containers should have the lid removed for at least 15 minutes every time you run the study, and sprayed if the soil is looking dry. Now that all of the containers have ventilation holes, leave them on the top shelf by the microwave so that they get more air circulation.

2.8.1. The bugs in the Tupperware containers are sow bugs, and the bugs in the container that looks like a legitimate terrarium are pill bugs. Please do not mix these bugs as that increases the likelihood that they will eat one another! Also, the sow bugs are going to be used as back up. Keep them happy and healthy in case we need them, but for the most part we will be using the pill bugs from the more official looking terrarium. Let me know if the bugs in any containers start to die so that I can order more if necessary.

2.9. If you have any bugs to run that day, get the empty Tupperware container and lay several sheets of damp paper towel neatly on the bottom. Using the brush and plastic spoon, gently place 25 bugs into this Tupperware container, seal the lid and then take it with you to LL1015J. Put it out of view for the participants (i.e. on top of the cabinet). Try to get as little dirt into the Tupperware as possible. If you notice that the main bug cage is getting dry, you can mist them with some water at this point. If you notice that the main bug cage has condensation on its sides/lid, it is too humid. Leave it with the lid off for the day as you conduct your studies, making sure to put it somewhere where it cannot get knocked over (on top of the book shelf over the computer monitors in the control room).

2.10. Put out the cups with either beans or bugs. To put bugs in the cups, use the brush again to sweep them into the cups. Try to keep dirt out of the cups, so that all they contain is either beans or bugs. Line the cups neatly in a 5x4 rectangle in front of the power outlets on the table, making sure that they aren’t too close to the edge, the power cord, or each other so that they will be easy for participants to pick up with one hand.
The bugs cannot crawl up smooth plastic surfaces, so they will be fine unattended in the cups for a few minutes. Leaving them in the cups for too long may cause them to dry out and die, however, so only place them in cups just before you are running a session.

2.10.1. Cover the cups with the box lid.

2.11. Log onto the computer in LL1015G, launch the survey and select the conditions to be run in the next session. Advance the survey to the consent page, shut the door.

2.12. Print out a copy of all of the SONA signups for that day. Put this in the RA’s clipboards. Keep a copy of this experimental protocol and a pen in the clipboard as well.

2.13. Make sure that you have enough copies of payment receipts printed and cut to size. Put these together with enough debriefing forms for that day in a neat pile in the control room.

2.14. Start up the camera in lab LL1015J & G and minimize that window but keep it running. Make sure that you do not accidentally start recording from the video camera.

3. **Running participants through the study**

3.1. Greet the participants in the waiting area outside the behavioural labs, 3-5 minutes before the session is due to start.

3.1.1. Approach people in the waiting room one at a time who look like they are waiting for a study, saying something like: “*Excuse me. Did you sign up for the Study of Physiological Changes During Behaviour at [state the time the study is due to start]?*” If yes, ask for the participants’ name and take their attendance. Lead them to LL1015G, instruct them to turn off any electronic devices and place everything in their bag on the floor out of the way. Instruct the participant to read through the consent form and electronically sign it if they don’t have any questions. When they have finished or if they have any questions, ask them to press the intercom button to call you.

3.2. As you wait for the participant to read the consent form, get the wrist monitor ready. When the participant buzzes the intercom, politely say something like: “*Hello. Have you finished? Okay, I’ll be right there.*” Bring the wrist monitor, stickers and tape with you.

3.3. When you enter the room to apply the wrist monitor, leave the door ajar so that no one can see in but the door is not closed (thus the participant will have privacy but not feel like they are trapped with you in the room). Explain to the participant that you are now going to set up the physiological measurement sensors. Say something like: “*Okay, so as mentioned in the consent form, I’m now going to help you put on some sensors to measure your pulse, skin perspiration, and blood flow. None of these sensors should be uncomfortable in any way as they simply sit on top of your skin and are held in*
place by stickers, tape and Velcro. [Tear off a 4-inch strip of tape and stick it to the desk so it is ready when you need it]. Which hand do you use to do most things, like write, throw a ball, brush your teeth etc? Okay, then we'll use your other arm.

First, I'll have you put on the main unit, which acts as a wireless transponder and heart monitor. Put it on just like a watch with the wires pointing towards your elbow and then fasten it snugly so that it's not too tight but doesn't move around on your arm. Okay next, I'll put on the skin thermometer which measures vasodilation by detecting changes in the temperature of your skin. This wire thermometer just wraps around your arm like so and then I hold it in place with a bit of tape. [Wrap the wire counter-clockwise around the arm of the participant and tape the end of the wire so that it doesn't flap around.] Lastly, I’m going to place these two stickers on your forth and fifth fingers. They contain a sensor and little bit of gel that helps them measure tiny amounts of skin perspiration more easily. Please press them firmly onto your fingers and don’t flex your fingers too much as they can sometimes fall off. Last, I just connect these wires to send the data back to the wrist transponder and we’re all set.

How’s that? Is anything bothering you or uncomfortable? [I doubt there will be any problems here, but fix participants’ complaints if they have any]. Okay, great. Now that we have the sensors in place we can start the study. Please follow me to the next room. You can leave your things here as you’ll be coming back to this room at the end of the study.

3.4. Lead the participants into LL1015J and ask them to stand in front of the middle of the table but not to touch any of the experimental materials yet. Say something like: “Please just stand still and relax while I turn on the receiver and make sure that it is properly calibrated. [Sit and then turn the receiver clock-wise so the back is clearly visible to participants and press the ‘on’ switch. – This will enable the participants to see that the equipment is indeed plugged in with the same cable that leads to the laptop. Turn the receiver unit back so that the wires face the wall and then say something like:]

“Okay, would you please turn on the transponder on your wrist by pushing the switch to the ‘on’ position?... Okay, great. The green light is on, that means the signal is good… Now let me see… [Look at the laptop monitor and open up the ‘simulated monitor animation’] Okay, great. The calibration seems to be good… everything is working just fine. Now let me just read you the official instructions for the task…” [Remove the lid from atop the cups and place it on the ground near your chair out of the way. Read the next part verbatim, and make it apparent that you are reading from a script].
3.4.1. **Bean condition – Flat rate (1):**

“You will now engage in a short action sequence using only your free hand (please don’t use the hand with the sensors as this can disrupt our measurement).

The purpose of this task is to measure your physiological response to payment incentives during a task that involves both fine and gross motor coordination, as well as auditory feedback. The payment you will receive for completing this task is $5. Therefore, when I say ‘begin,’ we need you to put as many of these beans as you can into the funnel in a 25 second period, one at a time, without spilling any. It is **very important** that you put as many as you can in the funnel during this period so that we can get a reliable measure your physiology during that 25 seconds. The beans will fall down the funnel to the blades of the grinding chamber of the coffee grinder. After the 25 seconds is up I will tell you to stop and then the grinder will be run for 5 seconds. The sound of the beans in the grinder will serve as the auditory feedback in this task. At the end of the study today, you will then be paid $5.

Unless you have any questions, we will start the task. Are you ready? **Begin** [start the stopwatch as you say begin] **Stop. Do not place any more beans into the funnel. Press the round ON button once to start the grinder now.”**

[When the participant presses the button, tell them that the grinder will automatically stop after 5 seconds. When the grinder shuts off, say something like the following as you lead them to LL1015G].

Okay, now that the action sequence is complete there are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.
3.4.2. **Bean condition – Piece rate (2):**

“You will now engage in a short action sequence using only your free hand (please don’t use the hand with the sensors as this can disrupt our measurement).

The purpose of this task is to measure your physiological response to payment incentives during a task that involves both fine and gross motor coordination, as well as auditory feedback. The payment you will receive for completing this task is $5 plus a performance bonus of 25 cents per bean that you place in the grinder. As there are 20 beans, you can earn up to a $5 performance bonus in addition to the guaranteed $5 payment for a potential total payment of $10.

Therefore, when I say ‘begin,’ we need you to put as many of these beans as you can into the funnel in a 25 second period, one at a time, without spilling any. It is very important that you put as many as you can in the funnel during this period so that we can get a reliable measure your physiology during that 25 seconds. The beans will fall down the funnel to the blades of the grinding chamber of the coffee grinder. After the 25 seconds is up I will tell you to stop and then the grinder will be run for 5 seconds. The sound of the beans in the grinder will serve as the auditory feedback in this task. At the end of the study today, you will then be paid $5 plus 25 cents per bean that you placed in the grinder during the next 25 seconds.

Unless you have any questions, we will start the task. Are you ready? Begin [start the stopwatch as you say begin] Stop. Do not place any more beans into the funnel. Press the round ON button once to start the grinder now.”

[When the participant presses the button, tell them that the grinder will automatically stop after 5 seconds. When the grinder shuts off, say something like the following as you lead them to LL1015G].

Okay, now that the action sequence is complete there are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.

3.4.3. **Bug condition – Flat rate (3):**

“You will now engage in a short task involving these roly-poly bugs. You will not have to touch the bugs, and for your information they are completely harmless as they do not bite, sting, or transmit any diseases to humans.

The task is a bug-extermination task in which you will use this extermination machine that we have made using a coffee grinder. Only use your free hand in the task (please don’t use the hand with the sensors as this can disrupt our measurement).

The purpose of this task is to measure your physiological response to payment incentives during a task that some people may consider somewhat unethical. The payment you will receive for completing this task is $5.

Therefore, when I say ‘begin,’ we need you to put as many of these bugs as you can into the funnel in a 25 second period, one at a time, without spilling any. It is very important that you put as many as you can in the funnel during
this period so that we can get a reliable measure your physiology during that 25 seconds. The bugs will fall down the funnel to the blades of the grinding chamber of the extermination machine. After the 25 seconds is up I will tell you to stop and then the grinder will be run for 5 seconds. At the end of the study today, you will then be paid $5.

Unless you have any questions we will start the task. Are you ready? Begin [start the stopwatch as you say begin] Stop. Do not place any more bugs into the funnel. Press the round ON button once to start the grinder now.”

[When the participant presses the button, tell them that the grinder will automatically stop after 5 seconds. When the grinder shuts off, say something like the following as you lead them to LL1015G].

Okay, now that the bug-extermination task is complete there are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.

3.4.4. If the participant refuses to do the task, say:

It is important for our research that you do this task.

If the participant refuses a second time, say:

Are you sure?

If the participant refuses a third time, say:

Okay. That’s fine. You will still receive your compensation for participating in this study. There are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.
3.4.5. **Bug condition - Piece rate (4):**

“You will now engage in a short task involving these roly-poly bugs. You will not have to touch the bugs, and for your information they are completely harmless as they do not bite, sting, or transmit any diseases to humans.

The task is a bug-extermination task in which you will use this extermination machine that we have made using a coffee grinder. Only use your free hand in the task (please don’t use the hand with the sensors as this can disrupt our measurement).

The purpose of this task is to measure your physiological response to payment incentives during a task that some people may consider somewhat unethical. The payment you will receive for completing this task is $5 plus a performance bonus of 25 cents per bug that you place in the grinder. As there are 20 bugs, you can earn up to a $5 performance bonus in addition to the guaranteed $5 payment for a potential total payment of $10.

Therefore, when I say ‘begin,’ we need you to put as many of these bugs as you can into the funnel, in a 25 second period, one at a time, without spilling any. It is very important that you put as many as you can in the funnel during this period so that we can get a reliable measure your physiology during that 25 seconds. The bugs will fall down the funnel to the blades of the grinding chamber of the extermination machine. After the 25 seconds is up I will tell you to stop and then the grinder will be run for 5 seconds. At the end of the study today, you will then be paid $5 plus 25 cents per bug that you placed in the grinder during the next 25 seconds.

Unless you have any questions, we will start the task. Are you ready?

*Begin.* [start the stopwatch as you say begin] *Stop.* Do not place any more bugs into the funnel. Press the round ON button once to start the grinder now.”

[When the participant presses the button, tell them that the grinder will automatically stop after 5 seconds. When the grinder shuts off, say something like the following as you lead them to LL1015G].

Okay, now that the bug-extermination task is complete there are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.

3.4.6. If the participant refuses to do the task, say word-for-word:

**It is important for our research that you do this task.**

If the participant refuses a second time, say word-for-word:

**Are you sure?**

If the participant refuses a second time, say:

Okay. That’s fine. You will still receive your compensation for participating in this study. There are a few questionnaires that we’d like you to complete on the computer where you read the consent form. You can leave the sensors on as the receiver works within a 60-foot range. When you’ve finished with the survey, just come back to this room and I’ll help you take the sensors off.
3.5. Once the participant has started the questionnaires on the computer, return to LL1015J and begin to set up for the next session. It’s okay if participants come into the lab while you are cleaning up as we are going to tell them at this point that no bugs were harmed during the experiment.

3.5.1. To empty the grinder, unscrew the grinder lid and then pour the funnel contents into the smaller Tupperware or a bowl. Reinstall the grinder lid and make sure that it works.

3.5.2. When the participant comes to find you, help them take the sensors off. Unclip the fingertip wires, peel off the tape, and then slide off the wrist monitor. Give them a debriefing sheet. If they were in the bug study, explain to them that no bugs were actually harmed in the study. Ask them, word-for-word: “did you know that the bugs were not killed in this study before I told you?” Also say something like: the study is being conducted to better understand people’s decision-making when they are asked to engage in unethical tasks. This research is important because people are often asked to engage in unethical tasks when at work: e.g. think of a business person who might be asked to lie in a negotiation on behalf of his or her company. The hope is that by better understanding how people make such decisions we can minimize the harm that organizations can sometimes do to others, society and the environment.

We hope that you understand why we had to use a little deception in this study. Obviously it’s very important for this study that other people who participate do not know about the study in advance. We hope that we can count on you not to talk about this study with others until the end of the summer.

3.5.3. Give the participant their money according to their condition and dismiss them. Have them sign and date a receipt for what they have earned and then put this receipt back into the cash box. Walk the participant out of the behavioural lab and close the door behind them.

4. When the participants have gone, set up the lab for the condition for the next session.

4.1. The last page of the survey is a fake last page that will progress to the real last page by clicking on the hidden button in the top right corner of the screen. On the actual last page fill out the data. Once you submit this, the program will tell you that the data have been recorded and then you can press alt+F4 to close the window. Relaunch the shortcut for the next study.

5. At the end of the day

5.1. Recycle SONA printouts in the recycling bins inside the control room.

5.2. Count your cash and enter the total into the same row in the ‘cash box accounting’ spreadsheet that you started the shift with. Send Julian an email telling him how much additional cash you will need by when for the next day that studies will be run. Place the cash box back in the filing cabinet with the biopac equipment. All of the equipment we
need should go on the top shelf to keep it separate from the other equipment. Put Chen-Bo’s laptop here as well.

5.3. Put the bugs back into their main Tupperware container. Does everything look okay in there? Notice anything that looks like mold/mildew? Keep an eye out for these bug killers. Mist them again if dry. Seal the main container. Throw the damp paper towel from the other container away and then leave this container with it’s lid askew.

5.4. Once we have finished collecting data, staple payment receipts to sheets of printer paper so that there are about 8 per sheet and all can be viewed by a scanner (i.e. make them look like the piece of paper from which they were cut). Keep these in the bottom of the cash box so that I can get this money back! Thanks!
Appendix B. Coding protocol for Experiment 1.

Instructions for Coding Participants’ Responses to Advertising Task

You and one other research assistant have been asked to code participants’ responses to an advertising task because you are not aware of the condition to which participants were assigned in the experiment. Therefore you are capable of coding participants’ responses more objectively than the researcher who designed and executed this study. Although your coding will be compared with that of the other research assistant in order to estimate reliability, please complete your coding independently of one another. If you have any questions about the coding scheme, please email both myself and the other research assistant and wait until the question is resolved to resume coding.

Participants in this study were asked to complete an advertising task. Please read the instructions provided to participants included at the bottom of this document so that you are familiar with what participants were asked to do.

The main coding task is to provide each message with a score on the basis of the following coding scheme. Enter your score in the corresponding column (T1Score, T2Score, etc.). If the participant did not enter a message for a particular tweet column, you may leave that score blank.

0 = Anti-tobacco or irrelevant or nonsensical message.

1 = Pro-tobacco message which does not exhibit high levels of creativity and/or persuasiveness. Messages given a score of 1 are not creative or persuasive, but they are relevant to the task and are not anti-tobacco.

2 = Pro-tobacco message which exhibits some creativity and/or persuasiveness. Give a message this score if you think it would potentially be useful in an advertising campaign on behalf of Philip Morris International or another tobacco company, provided they made some changes to the wording or content. Messages given a score of 2 would probably have required some thought on behalf of a participant who was trying to create a genuinely useful advertisement as part of the task. A message given a score of 2 does not have to be as high quality as professional advertising messages, but it should indicate that the participant was making a genuine effort to produce quality messaging.

3 = Pro-tobacco message which exhibits high levels of creativity or persuasiveness. Give a message this score if you think it is similar to the kinds of messages that a professional advertising campaign would use. Ask yourself if it could be paired with photos and logos to make a truly persuasive ad. A score of 3 doesn’t necessarily indicate that a message is perfect or among the best quality ads you have ever seen, but it should indicate that the participant was trying hard to produce high quality messaging that could be used by a tobacco company with
very few if any changes (not including any changes that might be required for compliance with regulations regarding the advertising of tobacco products).

The secondary coding task is to enter either a 1 or 0 into the column labelled “Did participant refuse?” Code this variable as -99 if the participant left all of the tweet columns blank. Code this variable as 0 if none of the messages the participant wrote are anti-tobacco (even if the messages are irrelevant to the task). Code this variable as 1 if any of the participant’s messages are anti-tobacco (either explicitly or implicitly – e.g. sarcastic or satirical or ironic messages that are intended to discourage tobacco use). Also code this variable as 1 if the participant uses the tweet space to express their unwillingness to write messages on behalf of Philip Morris / tobacco companies.

Once you have coded the messages written by the first 20 participants, email me your spreadsheets. This first 20 will constitute our calibration phase and I will assess the extent to which all of us agree on how this coding scheme applies to the messages written by participants.