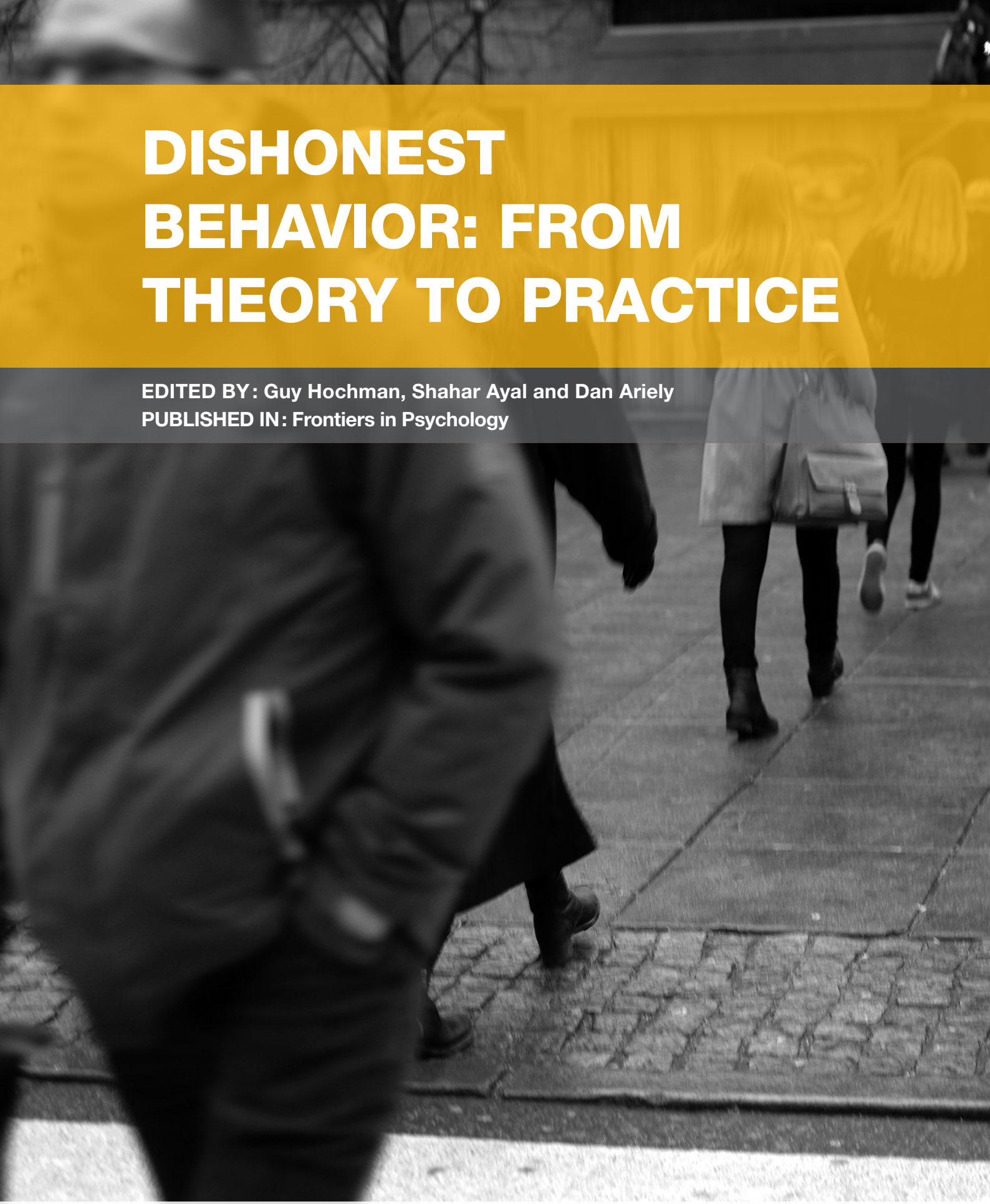

DISHONEST BEHAVIOR: FROM THEORY TO PRACTICE

EDITED BY: Guy Hochman, Shahar Ayal and Dan Ariely
PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88945-027-5
DOI 10.3389/978-2-88945-027-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

DISHONEST BEHAVIOR: FROM THEORY TO PRACTICE

Topic Editors:
Guy Hochman, Baruch Ivcher School of Psychology, Israel
Shahar Ayal, Baruch Ivcher School of Psychology, Israel
Dan Ariely, Duke University, USA

Image by Guy Hochman

The rapidly growing field of behavioral ethics shows that dishonest acts are highly prevalent in all walks of life, from corruption among politicians through flagrant cases of doping in sports, to everyday slips and misdemeanors of ordinary people who nevertheless perceive themselves as highly moral. When considered cumulatively, these seemingly innocuous and ordinary unethical behaviors cause considerable societal damage and add up to billions of dollars annually. Research in behavioral ethics has made tremendous advances in characterizing many contextual and social factors that promote or hinder dishonesty. These findings have prompted the development of interventions to curb dishonesty and to help individuals become more committed to ethical standards. The current e-book includes studies that test and advance current theory and deepen our understanding of the cognitive and physiological processes underlying dishonest behavior, discuss possible implications of findings in behavioral ethics research for real life situations, document dishonest behavior in the field and/or directly examines interventions to reduce it.

Citation: Hochman, G., Ayal, S., Ariely, D., eds. (2016). Dishonest Behavior: From Theory to Practice. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-027-5

Table of Contents

Section V – From theory to practice: Research using real-world situations and field data

frontiers in Psychology

EDITORIAL
published: 30 September 2016
doi: 10.3389/fpsyg.2016.01521

Editorial: Dishonest Behavior, from Theory to Practice

Shahar Ayal[1]*, *Guy Hochman*[1] and *Dan Ariely*[2]

[1] Interdisciplinary Center Herzliya, Baruch Ivcher School of Psychology, Herzliya, Israel, [2] Center for Advanced Hindsight, Social Science Research Institute, Duke University, Durham, NC, USA

Keywords: dishonesty, moral self, ethical dissonance, cross cultural, interventions

The Editorial on the Research Topic

Dishonest Behavior, from Theory to Practice

The rapidly growing field of behavioral ethics has shown that dishonest acts are highly prevalent in all walks of life, from corruption among politicians to flagrant cases of doping in sports, to everyday slips, and misdemeanors by ordinary people who nevertheless perceive themselves as highly moral. For instance, managers exaggerate travel expenses, consumers engage in wardrobing, citizens evade taxes, or download illegal music. When considered cumulatively, these seemingly innocuous and ordinary unethical behaviors cause considerable societal damage and add up to billions of dollars annually (Ariely, 2012).

Recent works in the behavioral ethics field have made tremendous advances in understanding the roots of dishonesty and characterizing the contextual and social factors that promote or hinder it. For example, one of the main insights is that people value morality and try to resist the temptation to act dishonestly (Aquino and Reed, 2002; Bazerman and Tenbrunsel, 2011). Investigations of misconduct in real life and in laboratory experiments indicate that while most people act dishonestly in everyday life, their dishonest acts are usually far below the maximum possible (Gneezy, 2005; Mazar et al., 2008; Shalvi et al., 2011). According to the Self-Maintenance model of dishonesty, this is due to *ethical dissonance* (Ayal and Gino, 2011; Barkan et al., 2012), a psychological tension which stems from the conflict between the desire to benefit from unethical behavior and the motivation to maintain a positive moral image (Barkan et al., 2015; Hochman et al., 2016).

The current research topic aims to utilize these lines of work to shift research in behavioral ethics from a descriptive approach to a more prescriptive and applicable one, thus advancing theoretical knowledge and making it possible to implement the findings to design and test practical interventions to promote ethical conduct among individuals in their day to day lives.

The first section explores the processes underlying dishonesty and highlights the interplay between moral self-image (MSI) and dishonesty. The second section sheds more light on contextual factors that promote or hinder dishonesty, with special attention to the perceived reasons and consequences of behavior. The last two sections emphasize the role of social and cultural norms both in the form of dishonesty as well as in effective interventions to reduce it.

MORAL SELF-IMAGE AND DISHONESTY

Several works examine the interplay between peoples' MSI and dishonest behavior. Jordan et al. aim to capture the fluctuations and malleability of the moral self in that people perceive themselves as highly moral, but engage routinely in unethical behavior. The paper defines the construct of MSI and presents an assessment questionnaire for moral self-perception in a current state. Liang et al. test how self-esteem affects the intention to engage in corrupt behavior. They show that increased self-esteem causes a low level of materialism, which in turn decreases corrupt intention.

OPEN ACCESS

Edited and reviewed by:
Eddy J. Davelaar,
Birkbeck, University of London, UK

***Correspondence:**
Shahar Ayal
s.ayal@idc.ac.il

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *18 August 2016*
Accepted: *20 September 2016*
Published: *30 September 2016*

Citation:
Ayal S, Hochman G and Ariely D (2016) Editorial: Dishonest Behavior, from Theory to Practice. Front. Psychol. 7:1521. doi: 10.3389/fpsyg.2016.01521

In a similar vein, Ding et al. delve deeper into the dynamics of MSI by examining the functions of guilt and moral identity in motivating prosocial behavior. They show that the link between acting immorally and compensatory behavior is mediated by guilt and moderated by moral identity (for related "moral accounting" models see Sachdeva et al., 2009; Gneezy et al., 2014). Merzel et al. show that people who acted dishonestly in the past are ready to suffer a future loss rather than admitting, even implicitly, that they lied. Chance et al. however show that this self-deception will decay if individuals are exposed to unbiased feedback about their true ability, but can be quickly revived if they get a new opportunity to cheat.

The last paper in this section (Burgoon) challenges the fundamental assumption that lying requires more effort than truth-telling. The paper discusses communication factors that may moderate the cognitive effort associated with producing deceptive messages.

CONTEXTUAL FACTORS THAT PROMOTE OR INHIBIT DISHONESTY

One way to resolve the tension between dishonest behavior and MSI is to creatively interpret an incriminating behavior as an honest or acceptable one (Mazar et al., 2008; Barkan et al., 2015). As a result, the magnitude of dishonesty is highly sensitive to contextual factors that affect our ability to justify unethicality (Shalvi et al., 2015; Hochman et al., 2016). Applying these insights, Rilke et al. show that self-reporting work hours dishonestly can be reduced by moving from a one-by-one to an all-at-once reporting policy. Van Der Zee et al. reported that negative emotional responses in an online settings (i.e., rejection) leads to increased dishonest behavior. By contrast, Lang et al. found that religious music can be used as a subtle cue associated with moral standards to curb dishonest behavior, but this mainly affects religious participants. Finally, the magnitude of dishonesty is also sensitive to the perceived identity of its victims. Amir et al. suggested that people are more willing to cheat groups than individuals, but only when the harm to the group is stated in global terms. In this context the lack of information about the harm caused to each individual can be used as a pretext for cheating.

CROSS-CULTURAL DIFFERENCES IN DISHONEST BEHAVIOR

Social and cultural norms play a key role in shaping moral behavior (e.g., Cialdini, 1993; Haidt and Joseph, 2004). In a study on cross-cultural differences in tax evasion between Italy and Sweden, Andrighetto et al. find that even though average tax compliance is similar in both countries, Italians were much more likely to fudge their income and Swedes were more likely to be completely honest or dishonest. Mann et al. found that legal sanctions and internal factors designed to deter minor, non-violent crimes have similar effects on different dishonest acts across five distinct cultures. More specifically, the results indicated that across countries, internal sanctions had the strongest deterrent effects on crime. However, the deterrent effects of legal sanctions were weaker and varied across countries.

FROM THEORY TO PRACTICE: RESEARCH USING REAL-WORLD SITUATIONS AND FIELD DATA

This research topic emphasized the applicability of unethical decision-making research in the real world. Moore and Pierce combines experimental and field data to examine how authorities penalize transgressors when the social context of the transgression elicits expectations of leniency. A surprising finding suggests that expectations of leniency (e.g., when the transgressor is caught on his birthday) appear to elicit psychological reactance and lead to stricter punishment.

The last two papers in this topic directly investigate the academic community itself, and speculate how to foster high ethical standards to improve scientific integrity. Rajah-Kanagasabai and Roberts show that engagement in research-related misconduct and questionable research practices is affected by attitudes, subjective, and descriptive norms about dishonesty, and mediated by justifications and behavioral intentions. Similarly, Schoenherr discusses potential practices (e.g., incentivizing quality rather than quantity of research) that may solve the problem of inappropriate authorship and encourage ethical behavior within the research community.

CONCLUSION

Research in the rapidly grown field of behavioral ethics suggests that public policies and interventions that are based on empirical research may encourage people to live according to higher ethical standards (Ariely, 2012; Ayal et al., 2015). The current research topic presents a wide array of research that contributes directly to this laudable goal. Taking together, these articles suggest that dishonest behavior in different forms and cultures share similar underlying processes. Thus, effective solutions to curb dishonesty and promote moral behavior in different domains (and across cultures) should be composed of the same psychological building blocks.

AUTHOR CONTRIBUTIONS

SA, GH, and DA equally contributed to the research topic and for the writing of this Editorial.

ACKNOWLEDGMENTS

The three guest editors would also like to personally thank all the authors and reviewers who contributed to this research topic. We also greatly appreciate the help of the Frontiers editorial office, as well as the kind assistance of Kiri Baga from the University of Pennsylvania.

REFERENCES

Aquino, K., and Reed, A. II. (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83:1423. doi: 10.1037/0022-3514.83.6.1423

Ariely, D. (2012). *The Honest Truth about Dishonesty*. New York, NY: HarperCollins.

Ayal, S., and Gino, F. (2011). "Honest rationales for dishonest behavior," in *The Social Psychology of Morality: Exploring the Causes of Good and Evil,* eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 149–166.

Ayal, S., Gino, F., Barkan, R., and Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Pers. Psychol. Sci.* 10, 738–741. doi: 10.1177/1745691615598512

Barkan, R., Ayal, S., and Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Curr. Opin. Psychol.* 6, 157–161. doi: 10.1016/j.copsyc.2015.08.001

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *J. Exp. Psychol.* 141, 757–773. doi: 10.1037/a0027588

Bazerman, M. H., and Tenbrunsel, A. E. (2011). *Blind Spots: Why We Fail to Do What's Right and What to Do About It*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400837991

Cialdini, R. B. (1993). *Influence: The Psychology* of *Persuasion*. New York, NY: Morrow.

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience accounting: emotion dynamics and social behavior. *Manage. Sci.* 60, 2645–2658. doi: 10.1287/mnsc.2014.1942

Haidt, J., and Joseph, C. (2004). Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus* 133, 55–66. doi: 10.1162/0011526042365555

Hochman, G., Fiedler, S., Glöckner, A., and Ayal, S. (2016). "I can see it in your eyes": biased processing and increased arousal in dishonest responses. *J. Behav. Decis. Mak.* 29, 322–335. doi: 10.1002/bdm.1932

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Sachdeva, S., Iliev, R., and Medin, D. L. (2009). Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychol. Sci.* 20, 523–528. doi: 10.1111/j.1467-9280.2009.02326.x

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications: doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 December 2015
doi: 10.3389/fpsyg.2015.01878

The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self

Jennifer Jordan[1], Marijke C. Leliveld[2] and Ann E. Tenbrunsel[3]*

[1] Department of Human Resource Management & Organizational Behaviour, University of Groningen, Groningen, Netherlands, [2] Department of Marketing, University of Groningen, Groningen, Netherlands, [3] Department of Management, University of Notre Dame, Notre Dame, IN, USA

Recent ethical decision-making models suggest that individuals' own view of their morality is malleable rather than static, responding to their (im)moral actions and reflections about the world around them. Yet no construct currently exists to represent the malleable state of a person's moral self-image (MSI). In this investigation, we define this construct, as well as develop a scale to measure it. Across five studies, we show that feedback about the moral self alters an individual's MSI as measured by our scale. We also find that the MSI is related to, but distinct from, related constructs, including moral identity, self-esteem, and moral disengagement. In Study 1, we administered the MSI scale and several other relevant scales to demonstrate convergent and discriminant validity. In Study 2, we examine the relationship between the MSI and one's ought versus ideal self. In Studies 3 and 4, we find that one's MSI is affected in the predicted directions by manipulated feedback about the moral self, including feedback related to social comparisons of moral behavior (Study 3) and feedback relative to one's own moral ideal (Study 4). Lastly, Study 5 provides evidence that the recall of one's moral or immoral behavior alters people's MSI in the predicted directions. Taken together, these studies suggest that the MSI is malleable and responds to individuals' moral and immoral actions in the outside world. As such, the MSI is an important variable to consider in the study of moral and immoral behavior.

Keywords: ethics, morality, self-image, self-concept, the self

OPEN ACCESS

Edited by:
Shahar Ayal, Interdisciplinary Center (IDC) Herzliya, Israel

Reviewed by:
Rachel Barkan, Ben-Gurion University of the Negev, Israel
Maryam Kouchaki, Northwestern University, USA

****Correspondence:***
Jennifer Jordan
j.jordan@rug.nl

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *23 June 2015*
Accepted: *20 November 2015*
Published: *15 December 2015*

Citation:
Jordan J, Leliveld MC and Tenbrunsel AE (2015) The Moral Self-Image Scale: Measuring and Understanding the Malleability of the Moral Self. Front. Psychol. 6:1878. doi: 10.3389/fpsyg.2015.01878

INTRODUCTION

Evidence of unethical behavior is widespread in society. From violations of psychological contracts (e.g., Kotter, 1973) to lying and deception (e.g., Lewicki, 1983), various forms of unethical behavior permeate modern life, creating both economic, and reputational costs. For many years, most empirical research on morality was dominated by the notion that there were stable, individual differences in moral behavior (e.g., Kohlberg, 1971; Colby et al., 1983; Kohlberg et al., 1983). However, contrary to the popular view that unethical behavior is just a matter of "a few bad apples," a rich body of recent literature demonstrates that even people who care about being moral (that is, those who have a high moral identity; Aquino and Reed, 2002) often engage in unethical behavior (for a recent review, see Bazerman and Gino, 2012). This research also argues that individuals' own perceptions of their morality is dynamic and malleable, and can influence subsequent behavior (Goldstone and Chin, 1993; Monin and Jordan, 2009; Shalvi et al., 2015): at any moment in time, social and situational factors may swing one's moral self-view. In the current investigation, we

propose the construct of the *moral self-image* (MSI), defined as a person's dynamic and malleable moral self-concept, to provide insight into this malleability of moral self-perceptions. We also propose a scale to measure the MSI. Across five studies, we demonstrate that this scale responds to feedback from the social world and people's reflections of their own moral behavior. By proposing the construct of the MSI, we hope to clarify how social and intrapersonal events, such as ethical and unethical behavior, shape people's views of their moral selves and how the state of their moral selves can affect their subsequent behaviors.

THE DYNAMIC AND MALLEABLE NATURE OF MORALITY

People engage in unethical actions on a daily basis, much more often than they care to admit (DePaulo et al., 1996; Schweitzer et al., 2004; Mazar et al., 2008; Gino et al., 2009; Shalvi et al., 2011). At the same time, they strive to maintain a positive self-concept both privately and publicly (Allport, 1955; Jones, 1973; Rosenberg, 1979; Adler, 2006). In fact, people wish to view themselves as moral beings (Steele, 1988; Dunning, 2007; Monin and Jordan, 2009) and take steps to maintain this belief when they behave immorally (Mazar et al., 2008; Monin and Jordan, 2009; Barkan et al., 2012; Shalvi et al., 2015)—even when these steps involve sacrificing gains or investing valuable resources (Murnighan et al., 1993; Dunning, 2007). According to recent research, when people act morally, their self-perception of their own morality is strengthened, allowing them to relax their subsequent moral strivings and engage in immoral actions. In contrast, after individuals act immorally, they seek to strengthen this self-concept by engaging in moral actions (Sachdeva et al., 2009; Jordan et al., 2011). Thus, the extent to which one's perceived morality "measures up" appears to be an important influence on actual (im)moral behavior.

This apparent discrepancy between people's perceived actual and ideal MSI leads to a dynamic and malleable perception of one's moral self: at any given moment, individuals may answer the question "How moral am I?" differently (Monin and Jordan, 2009; Moore and Gino, 2013). We label the answer to this question as a description of a person's MSI.

RESEARCH CALLING FOR THE MSI[1]

Within the rich body of research on the "self" (both the general and moral self), significant research proposes that the dynamics of the moral self explain immoral and moral behavior, yet no validated tool has been provided to measure this process (Zhong and Liljenquist, 2006; Sachdeva et al., 2009; Barkan et al., 2012; Mulder and Aquino, 2013).

For example, in their research on moral cleansing, Zhong and Liljenquist (2006) discuss the need for people to cleanse the moral self following an immoral act due to the need to self-complete via symbolic actions (Wicklund and Gollwitzer, 1981). The authors discuss that the need to do a good deed following a bad one is motivated by a desire for "restoration or completion of the moral self" (p. 1452); however, they do not measure the moral self nor provide empirical evidence that engaging in this type of deed actually affects people's moral selves.

In a similar vein, Mulder and Aquino (2013) demonstrate that people—particularly those with a high centrality of moral traits (i.e., a high internalized moral identity)—engage in behaviors that help to, "maintain their self-image as a moral person in the aftermath of a dishonest act" (p. 219). Mulder and Aquino find that following cheating, people who hold moral traits to high self-importance will engage in compensatory moral behavior. Although, they do not measure actual changes in one's self-image, they propose that this pattern is a consequence of a desire to "uphold a moral self-image" (p. 219) and reduce the discomfort of violating one's MSI.

Sachdeva et al. (2009) suggest that a person's need to boost the MSI (or what they term the "moral self-worth" and "moral self-concept") is responsible for compensatory patterns of moral behavior: "That is, when moral self-worth is threatened, moral cleansing restores the moral self-concept, and when moral self-worth is too high, moral licensing allows the agent to restrict moral behavior and return to a more comfortable level" (p. 524). Across a series of three studies, they find that when people write stories about themselves that affirm or threaten their MSI, they then act in opposing directions on subsequent tasks: a flattering story is followed by less moral behavior and an unflattering story is followed with more moral behavior. They also find that these effects do not occur if the story is written about someone else, suggesting that it is moral *self*-image that is at play, though this possibility is not empirically explored.

In a nuanced examination of the influences on dishonest behavior, Mazar et al. (2008) propose that people will be dishonest for self-gain—but only to the extent that dishonesty does not threaten their MSI. Mazar and colleagues use several paradigms in which people have the opportunity to cheat. Across five studies, they find that people *do* cheat—not always in a way that maximizes self-gain, but always in a way that, as they argue, protects their cherished MSI. For example, the researchers find that when dishonesty is framed in a way that makes a person mindful of her moral self-standards, she refrains from cheating in an effort to preserve her MSI. Mazar and colleagues argue that prior theories of dishonesty have failed to account for the value people place on maintaining their MSI, instead favoring a viewpoint that emphasizes a cost-benefit analysis on the part of the cheater. The researchers attempted to identify actual changes to participants' MSI by asking them about how moral they view themselves to be, but these questions yielded no effects.

We see a similar emphasis on MSI as an explanatory process in recent work that utilizes a cognitive dissonance framing to explain the effects of behavior on people's MSI (Shalvi et al., 2015). Barkan et al. (2012) demonstrated that having people contemplate their immoral misdeeds subsequently lowered their state self-esteem (Heatherton and Polivy, 1991; which then explained their greater willingness to punish and negatively judge other wrong-doers). Across six studies, they found that having people think about an unethical behavior that produced guilt, shame, or regret led participants' to report lower general self-images (compared to recalling neutral or favorable

[1]The literature has used other labels for what we refer to as the moral self-image, from moral self-concept (Mazar et al., 2008; Sachdeva et al., 2009) to moral self-worth (Sachdeva et al., 2009).

situations about the self). However, the authors also found that recalling a domain-general personal failure or an amoral behavior that elicited cognitive dissonance produced the same lowered state self-esteem. We suspect that had the authors specifically measured participants' MSI, changes would have only occurred in response to the immoral recall (see Studies 4 and 5 in the current manuscript for support for this supposition).

Lastly, Monin and Jordan (2009) discussed the emergence of a construct that captured the dynamics of the moral self. In their theoretical piece they discuss "a view of the self that is more reflective and more labile—one's moment to moment question of 'How moral am I?'" (p. 347), a question that they say people constantly strive to answer favorably. They explicitly call for a tool to measure the mechanism between an individual and his or her behavior, saying that understanding the dynamics of the moral self will "broad[en] the scope of phenomena that can be studied" (p. 348).

The literature reviewed above proposes the dynamics of the MSI as a mechanism for the dynamics of moral and immoral behavior. Yet because no tool is provided to empirically measure the state of the moral self and its dynamics, these assertions lack empirical evidence. One exception is a recent paper by Cornelissen et al. (2013), who found that when people were asked to recall a behavior they had performed that had a moral or immoral outcome, they compensated in their dishonesty—that is, they were more likely to cheat on a subsequent task (Mazar et al., 2008). This moral compensation was explained by differences in participants' MSI, which were measured using the scale proposed in the current investigation. More specifically, they used our scale[2] to demonstrate that the rise in MSI following a moral behavioral recall and the lowering of MSI following an immoral behavioral recall explained the magnitude of people's subsequent cheating behavior (i.e., more after moral behavior and less after immoral behavior). However, it is important to note that they did not demonstrate that (im)moral action recalls *changed* people's MSI from a baseline, a proposition that is central to our current theoretical argument. Our goal is to provide a theoretical foundation and empirically-driven examination of the MSI.

THE MSI AND THE SELF

We propose that the MSI resides in individuals' working self-concept, or current self-appraisal (Kernis and Johnson, 1990). The working self-concept is a malleable part of the self, which differentiates it from similar, more stable constructs, such as self-esteem (Rosenberg, 1965) and moral identity (Aquino and Reed, 2002). Like other areas of the working self-concept, people evaluate the state of their moral selves and attach either negative or positive labels to it based on cues from the social world and their own actions (Kernis and Goldman, 2003). Also like other parts of the working self-concept, the MSI is completely subjective, meaning that it is not a measure of the strength of one's moral judgments (Kohlberg, 1994), nor does it measure how moral (or immoral) a person actually is, but rather how moral (or immoral) she thinks she is. To take an extreme example, a devoutly religious person who dedicates his life to working with underprivileged children in the inner city might have a lower MSI following a spat with a fellow driver than a solipsistic investment banker who just made a small charitable donation following a similar argument. Though, individuals may vary in terms of how highly they value their moral selves, in general (see Aquino and Reed, 2002) people share a fairly universal desire to be moral (Dunning, 2007; Reed et al., 2007)—at least in terms of their self-perceptions of such morality (Mazar et al., 2008).

We define MSI as a person's malleable moral self-concept, that is, their self-concept related to the traits of the prototypically moral person (i.e., *caring, compassionate, helpful, hard-working, friendly, fair, generous, honest*, and *kind*)—derived from Aquino and Reed's (2002) work on the moral identity. While these nine traits are not expected to be an exhaustive representation of the traits of the moral prototype, we use these traits to evoke the mental representation of people's MSI. Below, we explain how the proposed construct of the MSI is associated with (and yet distinct from) other theoretically-related constructs.

Moral Identity

The MSI is distinct from moral identity in both its stability, as well as its implications for and responses to moral behaviors. Defined as "a self-conception organized around a set of moral traits" (Aquino and Reed, 2002, p. 1424), moral identity is comprised of an internalization subdimension, which is the importance to the self of possessing such traits, and a symbolization subdimension, which is the importance of demonstrating to others that one possesses those traits through one's behavior, style of dress, et cetera. Like the MSI, one's moral identity is conceptualized as a self-regulatory mechanism and is associated with various beliefs, attitudes, and behaviors (Aquino and Reed, 2002). But unlike the MSI, moral identity is a relatively stable trait (Aquino et al., 2009): "The definition of moral identity proposed here implies that if the identity is deeply linked to a person's self-conception, it tends to be relatively stable over time" (Aquino and Reed, 2002, p. 1425). If moral identity is highly regarded by the individual, it is predicted to lead to consistent moral actions throughout his or her life (Damon and Hart, 1992; Aquino and Reed, 2002). By contrast, the MSI is theorized to respond to events with a moral component, with a weak MSI stimulating moral action and a strong MSI allowing for moral relaxation (Cornelissen et al., 2013). Taken together, although the MSI is based on the moral traits identified by Aquino and Reed (2002), the MSI focuses on one's perception of how they are performing vis a vis these traits at a given moment, but not does not measure the extent to which a person values the moral traits (MI-internalization dimension) nor wishes to demonstrate them to others (MI-symbolization).

Self-Esteem

A vast amount of research exists on self-esteem (e.g., Deci and Ryan, 1995; Greenwald and Farnham, 2000; Crocker and Wolfe, 2001), which is defined as a person's global feelings of self-worth (Kernis and Goldman, 2003). Although, a person's self-esteem can change, it is unlikely to change in response to a single event

[2]Cornelissen and colleagues cited an earlier version of this manuscript, which had been presented at the Association for Consumer Research conference in St. Louis, MI in 2011: *Rules or Consequences? The Role of Ethical Mind-Sets in Moral Dynamics*.

or within a short period of time. Any instability in self-esteem usually occurs over an extended period (Rosenberg, 1986)—for example, from elementary to high school (e.g., McCarthy and Hoge, 1982). Distinctions have been made between global self-esteem and specific self-appraisals (similar to the MSI). Global self-esteem tends to be based on a generalized emotional response to the social world, whereas self-appraisals involve the cognitive appraisal of one's performance or acumen in some domain (Brown, 1993). While self-appraisals can influence self-esteem if they represent a core dimension of one's self-concept (Kernis et al., 1993; Pelham, 1995), global self-esteem and MSI differ from each other in three key ways: (1) self-esteem concerns a person's global feelings of self-worth rather than his or her specific moral self-appraisals, (2) self-esteem is relatively stable, and (3) self-esteem is more of an emotional than a cognitive response to the social world.

A more dynamic variation of self-esteem is state self-esteem, which is defined as a person's momentary assessment of self-regard. State self-esteem contains three sub-dimensions: performance (i.e., concern about one's abilities), social (i.e., concern about how others see oneself), and appearance (i.e., concern about how one physically appears; Heatherton and Polivy, 1991). Similar to our theorizing about the MSI, state self-esteem is affected by the environment. For example, Heatherton and Polivy (1991) find that state self-esteem decreases from a baseline following feedback about failure on an intellectual task and increases with interventions aimed at improving self-esteem. However, state self-esteem encompasses people's general feelings of self-worth rather than their specific feelings about their self-worth in the moral domain. Similar to the self-appraisal reasoning described above, we expect that while one's MSI would be likely to affect one's state self-esteem, the opposite is unlikely to be the case (i.e., one's general feeling of self-worth would not affect one's MSI).

Actual, Ought, and Ideal Selves

Self-discrepancy theory postulates that individuals have three "selves": the actual self, or the person one is perceived to be, the ideal self, or the person one would like to be, and the ought self, or the person one should be (Higgins, 1987). These latter two selves are referred to as the "self-guides," for they are thought to guide people's behavior and the nature of their self-assessments. Self-discrepancy theory contends that people are motivated to reach a state where their perceived actual self matches one of these self-guides. It also contends that discrepancies between what a person perceives to be his or her actual self and either the ideal or ought self lead to various types of negative emotions and discomfort. Despite apparent similarities, the construct of the MSI differs from self-discrepancy theory in two key ways. First, self-discrepancy theory concerns one's general self-assessment across domains rather than one's specific self-assessment in the moral domain (i.e., to measure one's self-discrepancy, people are asked to generate attributes related to each of the three selves). Second, self-discrepancy theory places significant importance on the source of the self view, proposing that each of these three selves are derived from either one's own self view *or* the individual's perception of how others perceive them (e.g., you have both the ought self that you perceive and an ought self that you think others perceive of you; see Higgins, 1987) and that the source of the self view is important because it affects the type of negative emotions or discomfort that results from discrepancies with the actual self. The MSI does not distinguish between the source of one's self-perceptions, as we contend that one's self-perceptions are a reflection of both one's own perceptions and the perceived perceptions of others. We also contend that the MSI is not exclusively derived from one of the two self-guides. Research on morality suggests that our moral standards come from a mix of the "oughts"—that is, the societally-dictated idea of what we should be (Kohlberg, 1971; Hoffman, 1975)—and "ideals," that is, the what we (or others) would like ourselves to possess (Lapsley and Lasky, 2001; Monin and Jordan, 2009; Jordan et al., 2011). While we believe that the MSI comes from a combination of the ideal and ought selves and assert that both the ideal and ought are influential self guides, we argue that one's personal ideal self-standards are more relevant to the MSI than are standards derived from the surrounding social context. Supporting this assertion is fundamental research on identity, which suggests that the self-concept is derived from the actual and ideal selves (Wylie, 1974), as well as research on self-esteem (a construct that we theorize and demonstrate is related to the MSI) which has been empirically associated with actual-ideal discrepancies but not actual-ought discrepancies (Moretti and Higgins, 1990). We address the distinction between the MSI and self-discrepancy theory both theoretically and empirically in Study 2.

THE CURRENT RESEARCH

Across five studies, we aim to formally present an instrument to measure the dynamics of people's moral selves, as well as to demonstrate its malleability in response to moral and immoral events. Study 1 demonstrates the convergent and discriminant validity of the MSI scale with other theoretically-related scales. Study 2 examines how the MSI is related to the *ideal* versus the *ought* self (Higgins, 1987). Studies 3 through 5 demonstrate the construct validity of the MSI, by demonstrating the malleability of this measure based on various events (i.e., feedback-related and self-related recalls). Study 3 looks at feedback related to social comparisons of moral behavior, and Study 4 examines feedback relative to one's own moral ideal. Lastly, Study 5 provides evidence that recall of one's moral or immoral behavior affects subsequent immoral behavior (Cornelissen et al., 2013), altering MSI in the predicted directions[3].

STUDY 1

Across two samples (1a and 1b), we compare our MSI measure with measures of theoretically-related constructs, including Moral Identity (Aquino and Reed, 2002), Generalized Self-esteem (Rosenberg, 1965), Moral Disengagement (Moore

[3]The procedures for all five studies in this manuscript received approval from an institutional review board prior to data collection. All procedures complied with the rules regarding conducting research with human subjects proposed by the American Psychological Association.

et al., 2012), Religiosity (Brown, 1962), Negative Reciprocity Norm (Eisenberger et al., 2004), and Sympathy (Ahmed and Jackson, 1979). We explain each of these constructs and their accompanying scales, as well as our hypothesized relationships with these constructs below.

Moral Self-Image

We measured MSI by presenting nine traits perceived as prototypical of the ideally-moral person (Aquino and Reed, 2002). Using a nine-point Likert Scale (1 = *much less than the X person I want to be*; 9 = *much more than the X person I want to be*), we asked people to indicate where they were relative to their ideal self on each trait; see Supplementary Material.

Moral Identity

Moral identity is defined as having a self-conception organized around a set of moral traits. Moral identity possesses two dimensions, *internalization* and *symbolization*. Internalization is the importance people place on possessing these traits, and symbolism is the importance they place on demonstrating these traits to others (e.g., through membership in clubs or the clothes they wear). For example, an *internalization* item is, "It would make me feel good to be a person who has these characteristics," whereas a *symbolization* item is, "I am actively involved in activities that communicate to others that I have these characteristics." Using the Aquino and Reed (2002) 10-item scale (1 = *completely disagree*; 7 = *completely agree*), we measured both dimensions and had divergent predictions for each. Previous research has demonstrated that past moral actions affect the symbolic but not the internalized moral identity (Jordan et al., 2011) and that, instead, the internalized moral identity affects how people behaviorally respond to immoral events (Mulder and Aquino, 2013). Thus, we hypothesized that while one's MSI would not be affected by the importance one places on possessing moral traits (internalization), it would be (positively) affected by the extent to which one demonstrates the moral self to others (symbolization), as such public demonstrations would boost people's conceptions of their moral selves.

Generalized Self-Esteem

Generalized self-esteem is defined as a person's global feelings of self-worth and -acceptance. We measured self-esteem using Rosenberg's (1965) 10-item measure. Items included, "On a whole, I am satisfied with myself," and "At times, I think I am no good at all (reverse-scored)" (1 = *strongly disagree*; 5 = *strongly agree*). Although, self-esteem is considered a stable construct and MSI is considered a malleable construct, we predicted a positive relationship between MSI and generalized self-esteem, given that temporary self-appraisals have been found to be predictive of global self-esteem, particularly when these self-appraisals are a part of the self that the person considers central or core (see Kernis et al., 1993; Pelham, 1995).

Moral Disengagement

Moral disengagement is defined as, "an individual's propensity to evoke cognitions which restructure one's actions to appear less harmful, minimize one's understanding of responsibility for one's actions, or attenuate the perceptions of the distress one causes to others" (Moore, 2008, p. 129). In other words, moral disengagement is a person's ability to rationalize his or her immoral behavior in a way that helps reduce the negative feelings that would otherwise result. We measured moral disengagement using the eight-item Propensity to Morally Disengage Scale (Moore et al., 2012), which included, "People shouldn't be blamed for doing things that are technically wrong when all their friends are doing it too," and "Some people have to be treated roughly because they lack feelings that can be hurt" (1 = *not at all*; 7 = *very much so*). We predicted that moral disengagement would be positively related to MSI because the more one is able to morally disengage from one's immoral actions, the greater one's MSI.

Religiosity

Religiosity is defined as the extent to which a person holds various religious beliefs. We measured religiosity using the Other Orthodox Christian Beliefs subscale of Brown's (1962) Religiosity measure. Intuitively, religiosity may be related to the perception of oneself as moral; indeed, religiosity and people's desire to symbolize their moral self to others have been found to be positively correlated (Aquino and Reed, 2002). Thus, we hypothesized a positive relationship between religiosity and MSI.

Negative Reciprocity Norm

The negative reciprocity norm is the belief that it is appropriate to retaliate against an immoral or unjust act leveled against oneself (Gouldner, 1960). We measured this construct using the nine-item scale of Eisenberger et al. (2004), which included, "If someone says something nasty to you, you should say something nasty back" and "If someone treats you badly, you should treat that person badly in return" (1 = *not at all*; 7 = *very much so*). The negative reciprocity norm has been found to be negatively related to the extent to which a person considers moral traits central to his or her self-concept (Aquino and Reed, 2002). As the MSI focuses on an assessment of the state of one's moral-self rather than how much one values a moral identity (which would likely be associated with someone's desire to retaliate for an act perceived as unjust), we did not expect any relationship between MSI and holding a norm of negative reciprocity.

Sympathy

Sympathy is the ability to show concern for the needs and welfare of others (Eisenberg, 2000). We measured sympathy with the eight-item nurturance dimension of the Acceptance of Welfare scale (Ahmed and Jackson, 1979), which included, "Someone who is disabled will get my attention and aid" and "People in need deserve my sympathy and support" (1 = *strongly disagree*; 7 = *strongly agree*). Similar to the rationale behind our predictions for Negative Reciprocity Norm, we predicted a null relationship between MSI and Sympathy. The state of one's MSI should be unrelated to one's general beliefs about showing sympathy for those less fortunate.

Positive and Negative Affect

In Sample 1b only we examined positive and negative affect because we wished to see how the state of the MSI related to individuals' affective states. We administered the PANAS (Watson et al., 1988), which presented participants with 10 positive (e.g., *proud, active*) and 10 negative (e.g., *upset, nervous*) items and asked them to rate themselves on each item based on how they were feeling at the current moment (1 = *not at all*; 7 = *very much so*). As the PANAS measures a state-based construct (similar to the MSI), and because people's moral selves are an integral part of their self-concepts, we predicted that the MSI would be positively related to positive affect and negatively related to negative affect.

Gender and Age

While there is some evidence that women reason differently (Jaffee and Hyde, 2000) and perhaps more complexly about moral issues than men (Wark and Krebs, 1996; White, 1999), there is no evidence to suggest that women think of themselves as any more or less moral then men. Similarly, there is evidence that moral behavior changes from adolescence into adulthood but is fairly stable in adulthood (Eisenberg et al., 2005), the age category of our samples. Thus, we predicted null relationships between MSI and both the demographic variables of age and gender.

Participants—Sample 1a

Participants were 574 American adults from a Mechanical Turk (Mturk) sample (M_{age} = 32.89, SD = 11.04, 48% female). They were invited to take part in a 20-min study in exchange for $0.55. Thirty participants did not pass the attention checks and thus were eliminated from the analyses, leaving a final sample of 544 on which to run the analyses.

Participants—Sample 1b

Participants were 515 American adults from an Mturk sample (M_{age} = 31.88, SD = 8.57, 49% female). They were invited to take part in a 20-min study in exchange for $0.60. Sixteen participants did not pass the attention checks and thus were eliminated from the analyses, leaving a final sample of 499 on which to run the analyses.

Procedures

Participants read a consent form and, if they agreed to the terms, logged on to the study website. They completed all measures (with the NRN and Sympathy scales only in Sample 1a and the PANAs only in Sample 1b) stated above in a randomized order; however, either the MSI scale or the moral identity scale always came first.

Results and Discussion

All results (including Cronbach Alphas for the measures) are contained in **Tables 1**, **2**.

Demonstrating convergent validity, across both samples, MSI was positively related to symbolic (but not internalized) moral identity, generalized self-esteem, moral disengagement, and religiosity; also as predicted, demonstrating divergent validity, we found no relationship between MSI and negative reciprocity norms and sympathy. However, in contrast to predictions, we found that in Sample 1a (but not 1b) age was positively related to MSI, with older individuals having higher MSIs than younger individuals. In Sample 1a gender was marginally negatively related to MSI, with women reporting higher MSIs than men; however in Sample 1b the directionality of this relationship flipped such that men reported higher MSIs than women.

We then conducted an exploratory factor analysis to explore the factor structure of our MSI scale, predicting that a single factor would emerge from the data. We conducted a principal components analysis with an oblique rotation method, which would allow for the potential factors to be correlated with one another (*direct oblimin*). From an inspection of the scree plot, eigenvalues, and factor loadings across both samples, only one factor (Sample 1a: Eigenvalue = 4.48; Sample 1b: Eigenvalue = 4.68) emerged from the data. This factor explained between 51.96% (1a) and 52.37% (1b) of the variance, and all items loaded on to this factor at a loading of 0.53 (*How hardworking*

TABLE 1 | Study 1 Sample 1a—Scale intercorrelations and reliabilities.

	MSI	MIs	MIi	GSE	MD	RELIG	NRN	SYMP	Age	Sex
MSI	0.88									
MIs	0.26***	0.83								
MIi	0.03	0.29***	0.84							
GSE	0.20***	0.17***	0.26***	0.93						
MD	0.15***	0.001	−0.44***	−0.18***	0.84					
RELIG	0.17***	0.25***	0.14***	0.11**	0.009	0.87				
NRN	0.05	−0.19***	−0.37**	−0.14***	0.55***	−0.09*	0.86			
SYMP	0.03	0.26***	0.63**	0.26***	−0.51***	0.12**	−0.48***	0.95		
Age	0.09*	−0.02	0.16***	0.21***	−0.22***	0.14***	−0.12**	0.13**	–	
Sex	−0.06†	−0.11**	−0.23***	−0.02	0.25***	−0.13**	0.19***	−0.17***	−0.17***	–

*Cronbach alphas contained in the diagonals. †$p < 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; MD, moral disengagement; RELIG, religiosity; NRN, negative reciprocity norm; SYMP, sympathy. For gender, 1, female; 2, male.*

TABLE 2 | Study 1 Sample 1b—Scale intercorrelations and reliabilities.

	MSI	MIs	MIi	GSE	MD	RELIG	PANAS-P	PANAS-N	Age	Gender
MSI	0.88									
MIs	0.23***	0.85								
MIi	−0.07	0.38***	0.82							
GSE	0.31***	0.23***	0.22***	0.94						
MD	0.15***	−0.05	−0.46***	−0.16**	0.84					
RELIG	0.14**	0.30***	0.25***	0.13**	−0.03	0.90				
PANAS-P	0.32***	0.36***	0.17***	0.50***	0.02	0.27***	0.92			
PANAS-N	−0.01	−0.08†	−0.31***	−0.39***	0.32***	−0.02	−0.04	0.92		
Age	0.06	0.00	0.11*	0.18***	−0.20***	0.14**	0.16***	−0.13**	–	
Gender	0.09*	−0.10*	−0.22***	0.04	0.18***	−0.14**	0.03	0.05	−0.14**	–

*Cronbach alphas contained in the diagonals. †$p < 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; MD, moral disengagement; RELIG, religiosity; PANAS-P, PANAS positive affect; PANAS-N, PANAS negative affect. For gender, 1, female; 2, male.*

are you relative to your ideal?) or higher[4]. In sum, across Samples 1a and 1b, we found that MSI was positively related to symbolic moral identity, generalized self-esteem, and religiosity and negatively related to moral disengagement. We also found that our scale contained a single factor structure, which explained at least 50% of the variance across both studies. These findings provide suggestive evidence of the validity of MSI as a unique construct.

STUDY 2

In Study 2, we examine how the MSI is related to the ideal versus ought self (Higgins, 1987). As noted earlier in the Introduction, self-discrepancy theory postulates that individuals have the actual self and two "self-guides," including the ought self and the ideal self (Higgins, 1987). Self-discrepancy theory argues that people are motivated to align their perceived actual self with one of these self-guides and that discrepancies between the actual self and either the *ought* or *ideal* self lead to negative emotions and discomfort. As argued above, we propose that the MSI is primarily comprised of one's perceived moral self relative to one's own moral ideal self-standard rather than relative to an externally-imposed standard (i.e., the ought). That is, the MSI assesses who a person perceives him or herself to be relative to the ideal moral person that he or she wishes to be—not the moral person he or she thinks *others* wish him or herself to be. This contention is derived from research demonstrating that moral or immoral behaviors do not need to be witnessed by others in order to elicit compensatory effects; only the individual him or herself needs to be aware of the event (Sachdeva et al., 2009; Jordan et al., 2011), as well as research suggesting that the self concept is comprised of a mix of actual and ideal states (Wylie, 1974).

In order to empirically test this idea, half of the participants in the current study completed the MSI scale as it was originally written (i.e., in a way that measured the ideal moral self). The other half of participants completed a version of the scale in which we asked people not about the moral self they perceived themselves to possess relative to where they wanted to be (*ideal*) but rather about the moral self they perceived themselves to possess relative to what they thought *others* wanted them to possess (*ought)*. Along with one of these two versions of the scale, participants also completed the same measures administered to Sample 1b in Study 1.

Although, we contend that the *ought* self is relevant for the MSI, we predicted that the ideal MSI would be a better fit than the *ought* moral self with our proposed model.

Participants, Design, and Procedures

Participants were 590 American adults from an Mturk sample (M_{age} = 35.94, SD = 11.34, 50.5% female). Participants were invited to take part in a 15-min study in exchange for $0.75 compensation.

[4]We also ran CFAs across Samples 1a and 1b. Specifically, to examine the veracity of our proposed model related to other plausible models, we used the confirmatory factor analysis function of LISREL 8.80 maximum likelihood estimation method. Model fit was assessed by the Root Mean Square Error of Approximation (RMSEA), the Normed-fit Index (NFI), and the Comparative Fit Index (CFI). In addition, competing models were compared to our proposed model by means of chi-square differences. For Sample 1a, in which we compared our proposed model in which MSI, the symbolic and internalized sub-dimensions of moral identity, general self-esteem, moral disengagement, sympathy, negative reciprocity norms, and religiosity were analyzed as distinct factors with three other models in which symbolic moral identity and generalized self-esteem, and moral disengagement were loaded on to the same factor as MSI. The eight-factor model demonstrated better fit [$\chi^2(1924)$=5519.58; RMSEA = 0.059 (0.057, 0.060); NFI = 0.93; CFI = 0.96] than all three other models [7-factor: $\Delta\chi^2(7) = 1721.13$; RMSEA = 0.071 (0.069, 0.073); NFI = 0.92; CFI = 0.95; 6-factor: $\Delta\chi^2(13) = 10{,}878.27$; RMSEA = 0.12 (0.12, 0.12); NFI = 0.88; CFI = 0.91; 5-factor: $\Delta\chi^2(18) = 10{,}847.16$; RMSEA = 0.12 (0.12, 0.12); NFI = 0.87; CFI = 0.90]. Similarly, for Sample 1b we compared an 8-factor model, which examined the MSI, the symbolic and internalized sub-dimensions of moral identity, generalized self-esteem, moral disengagement, religiosity, and positive and negative affect as distinct factors with five other models in which the MSI was loaded on to factor along with symbolic moral identity, generalized self-esteem, and positive affect. The 8-factor model demonstrated better fit [$\chi^2(1741) = 4588.07$; RMSEA = 0.057 (0.055, 0.059); NFI = 0.95; CFI = 0.95] than all comparison models [7-factor with MSI and symbolic MI loaded together: $\Delta\chi^2(7) = 1769.47$; RMSEA = 0.073 (0.071, 0.075); NFI = 0.90; CFI = 0.93; 7-factor with MSI and generalized self-esteem loaded together:$\Delta\chi^2(7) = 3774.63$; RMSEA = 0.087 (0.085, 0.089); NFI = 0.89; CFI = 0.92; 7-factor model with MSI and positive affect loaded together:$\Delta\chi^2(7) = 3417.13$; RMSEA = 0.085 (0.083, 0.087); NFI = 0.92; CFI = 0.92; 6-factor model: $\Delta\chi^2(13) = 5721.57$; RMSEA = 0.99 (0.97, 0.10); NFI = 0.89; CFI = 0.90; 5-factor model: $\Delta\chi^2(18) = 10778.19$; RMSEA = 0.12 (0.12, 0.13); NFI = 0.86; CFI = 0.86].

TABLE 3 | Study 2—Scale intercorrelations and reliabilities for the *Ideal* moral self.

	MSI	MIs	MIi	GSE	MD	RELIG	PANAS-P	PANAS-N	Age	Gender
MSI	0.88									
MIs	0.34***	0.84								
MIi	0.05	0.36***	0.85							
GSE	0.20***	0.21***	0.15**	0.93						
MD	0.14*	−0.02	−0.46***	−0.11†	0.81					
RELIG	0.17**	0.30***	0.30***	0.08	−0.10†	0.89				
PANAS-P	0.33***	0.30***	0.25***	0.42***	−0.01	0.24***	0.92			
PANAS-N	−0.03	0.05	−0.16**	−0.26***	0.31***	0.02	−0.11	0.88		
Age	−0.03	0.004	0.15**	0.08***	−0.19***	0.20***	0.15**	−0.08	–	
Gender	−0.03	−0.21***	−0.18**	−0.06	0.20***	−0.18**	0.002	−0.03	−0.08	–

*Cronbach alphas contained in the diagonals. †$p < 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI,moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; MD, moral disengagement; RELIG, religiosity; PANAS-P, PANAS positive affect; PANAS-N, PANAS negative affect. For gender, 1, female; 2, male.*

Participants were randomly assigned to complete either the MSI scale or a version of the MSI scale in which we asked about their *ought* moral selves. Specifically, in the *ought self* condition, instead of asking participants to indicate how *caring, compassionate*, *fair*, et cetera he or she was at the present time relative to the person who he or she wanted to be, we phrased these items so that the participant was asked to indicate how *caring, compassionate, fair*, et cetera he or she was at the present time relative to who *others* wanted him or her to be. All participants then completed measures of moral identity (Aquino and Reed, 2002), generalized self-esteem (Rosenberg, 1965), moral disengagement (Moore, 2008; Moore et al., 2012), religiosity (Brown, 1962), and positive and negative affect (Watson et al., 1988).

Results and Discussion

All results (including Cronbach Alphas for the measures) are contained in **Tables 3, 4**.

For the MSI (i.e., the *ideal* moral self), all relationships found in Study 1 (except for the relationship with gender) were replicated in the current study. Specifically, demonstrating the convergent validity of the proposed construct, MSI was positively related to symbolic (but not internalized) moral identity, generalized self-esteem, moral disengagement, religiosity, and positive affect. And again, demonstrating the divergent validity of the MSI with other constructs, we found no relationship between the MSI and negative affect or age. In this study, there was no relationship with gender.

In contrast, while several of the relationships found between MSI and the other explored constructs were replicated when using the *ought* version of the MSI scale, unlike the *ideal* MSI, the *ought* version showed a moderate positive correlation with the internalization subdimension of moral identity (Aquino and Reed, 2002), no correlation with moral disengagement (Moore, 2008), and a positive correlation with generalized self-esteem (Rosenberg, 1965), which was double the magnitude as witnessed for the *ideal* moral self. We also saw a moderate-sized correlation with gender, such that women reported having a greater moral self as perceived by others.

Thus, it appears that except for the negative correlation with negative affect, the *ideal* MSI more accurately captured our hypothesized relationships with the predicted related (and unrelated) constructs. To empirically test this assertion, we used the confirmatory factor analysis function of LISREL 8.80 maximum likelihood estimation method to compare our purported model using both the *ideal* MSI and the *ought* moral self via assessing the chi-square differences between the two models. As done in the previous studies, model fit was assessed by the Root Mean Square Error of Approximation (RMSEA), the Normed-fit Index (NFI), and the Comparative Fit Index (CFI).

The first model tested was the purported model in which the (*ideal*) MSI, the symbolization and internalization subdimensions of moral identity, generalized self-esteem, moral disengagement, religiosity, and positive and negative affect were analyzed as distinct factors. When using the *ideal* MSI, this eight-factor model had a good fit with the data, $\chi^2(1801) = 3917.94$; RMSEA = 0.064 (0.061, 0.066); NFI = 0.87; CFI = 0.93. In contrast, while the model using the *ought* MSI also showed sufficient model fit, $\chi^2(1801) = 3898.68$; RMSEA = 0.063 (0.060, 0.065); NFI = 0.90; CFI = 0.95, it was inferior to the one using the *ideal* MSI, $\Delta\chi^2(1) = 19.26$, $p = 0.00001$[5].

In sum, while we did find that the *ought* moral self showed many of the same relationships as were found with the *ideal* MSI, the *ought* moral self was strongly positively correlated with individuals' internalized moral identity, which is the more "trait-like," stable dimension of the two moral identity subdimensions (Jordan et al., 2011). It was also strongly positively correlated with the generalized self-esteem—a stable personality dimension. Taken together, it appears that the *ought* moral self mimics more of a stable, individual difference than does the *ideal* MSI. As stated earlier, we see the MSI not as being a stable, individual difference but as a state that responds to people's moral actions and social comparisons to the world around them. More research is required to make statements about the stability of the *ought* moral self with confidence.

[5]Although, the degrees of freedom were equivalent for both models, you cannot test the significance of a chi-square value with a degrees of freedom equal to 0. Thus, we set this to "1," which is a conservative test of our hypothesis.

TABLE 4 | Study 2—Scale intercorrelations and reliabilities for the *Ought* moral self.

	MSI	MIs	MIi	GSE	MD	RELIG	PANAS-P	PANAS-N	Age	Gender
MSI	0.91									
MIs	0.42***	0.87								
MIi	**0.41***	0.33***	0.87							
GSE	**0.40***	0.26***	0.25***	0.94						
MD	**-0.10**	0.05	−0.48***	−0.24***	0.84					
RELIG	0.22***	0.23***	0.018**	0.11†	0.04	0.89				
PANAS-P	0.39***	0.37***	0.21***	0.42***	0.03	0.22***	0.93			
PANAS-N	−0.16**	−0.06	−0.38***	−0.36***	0.42***	0.01	0.02	0.90		
Age	0.08	−0.04	0.14**	0.04	−0.16**	0.14**	0.05	−0.10	–	
Gender	**−0.19****	−0.18**	−0.30***	−0.09	0.23***	−0.30***	−0.05	0.08	−0.13*	–

*Cronbach alphas contained in the diagonals. Correlations that are bolded are those in which the relationship differed between the ought and the ideal moral self. †$p < 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; MD, moral disengagement; RELIG, religiosity; PANAS-P, PANAS positive affect; PANAS-N, PANAS negative affect. For gender, 1, female; 2, male.*

Unlike the *ideal* MSI, the *ought* moral self was negatively correlated with negative affect. This is a relationship that was predicted for the *ideal* moral self but not witnessed in either Studies 1b or Study 2. Why this is is unknown. Perhaps, as highly social beings (Aronson, 2003), thinking about how others see you more strongly elicits negative emotions than does thinking about one's own moral self-evaluation. And lastly, we found that the *ought* moral self was higher for females than for males. While this was not hypothesized (nor found) for the *ideal* MSI, it was found for the *ideal* moral self in Study 1's Sample 1a. Why it was not found in the current study is unknown. However, it is unsurprising that women reported having a higher moral self as conceived by others in their social world, as there is evidence that society views women as being more moral and virtuous than men (Fiske et al., 1999, 2002; White, 1999).

In Studies 3, 4, and 5 we examine a fundamental assertion that underlies our theorizing about MSI, namely that the MSI responds to explicit feedback about one's moral state relative to others and to one's own moral ideal. Research on the self demonstrates that the self-concept is influenced by three primary sources originating in the social environment: social comparison, feedback, and an individual's actions (Kernis and Goldman, 2003). In Study 3, we explore the effect of the first of these three sources, social comparison, on changes to a person's MSI (looking specifically at feedback in Study 4 and actions in Study 5). In Studies 4 and 5, in order to examine the independent contribution of MSI, we then investigate whether such feedback influences related constructs. Specifically, we investigate whether such feedback not only affects the MSI but also moral identity (Aquino and Reed, 2002), generalized self-esteem (Rosenberg, 1965), and state self-esteem (Heatherton and Polivy, 1991).

STUDY 3

In Study 3, we explore one of the primary sources of influence originating in the social environment (Kernis and Goldman, 2003), social comparison information, and examine how it influences the MSI. We predicted that the MSI would be affected by this feedback with positive feedback leading to an increase in one's MSI and negative feedback leading to a decrease.

Participants and Design

Participants were 59 international business students (56% women, $M_{age} = 21.86$, $SD = 2.45$) at a university in the Netherlands who participated in exchange for €6. We presented all materials in English and randomly assigned participants to one of two moral-valence conditions: above average moral or below average moral. Thirteen participants were excluded from the analyses, leaving us with a total of 46 participants on which to run the analyses[6].

Procedures

Participants were required to complete our MSI scale at least 15 h prior to coming into the lab. We sent them a link to the MSI scale immediately after they had signed up for the study.

When the participants arrived in the lab, we told them they would be completing a study about their environmental conservation behavior. Participants were required to write a short essay about, "what actions you take in support of environmental conservation and why you think these are important." We used the topic of environmental conservation behavior because previous research has found this topic to be related to people's moral selves (Mazar and Zhong, 2010). We told each participant that the experimenter would interrupt him or her after several minutes to obtain more information about the essay he or she had just written. Before the experimenter came in, the computer delivered a message to the participant. They were told that the essay they had written was actually part of a standardized measure of people's "MIP," or "how much moral traits are a part of your identity and who you are." We said that the measure assessed both the vocabulary they used and the speed at which they typed to generate a score that we could compare with the

[6]Six participants were excluded due to behavior during the lab session (e.g., could not understand English or a fire alarm occurred in the middle of the session, forcing the lab to be evacuated), three because they did not believe that the MIP was a real test, and four because they took the post-test but did not take the pre-test.

scores of others in the population. The experimenter then opened the door and gave participants a sheet that further explained this measure and the ranges of scores that were possible; these ranges were presented in five categories ranging from a very low moral self-identity to a very high one. The experimenter told participants that he would type a personal code into the main computer that would allow the participant to see his or her score. He assured each participant that this score would only be visible to the participant. Following this interaction, the participant saw his or her score. This score fell into one of two categories: very high or very low relative to the rest of the population. Each score was accompanied by the percentiles of the population in which they fell (e.g., 1st–11th percentile; 88–99th percentile), hence providing a point of social comparison. The participant then completed the MSI scale once again.

Before leaving the lab, we asked participants to indicate the range their score fell into from a choice of five options. Finally, they were fully debriefed, a process that included telling them that the measure and associated feedback were completely bogus.

Results

Manipulation Checks

All participants selected the correct score range on the manipulation check.

MSI

In order to analyze our hypothesis that feedback would be directly related to a change in individuals' MSI, we subtracted their score on the pre-test from their score on the post-test (for similar methods, see Heatherton and Polivy, 1991). In a case like this, where a change score is used as the dependent, rather than independent, variable, polynomial regression is not necessary nor appropriate (see Edwards, 2002)[7].

Participants in both the extremely positive ($M = -0.25$, $SD = 0.76$) and extremely negative ($M = 0.22$, $SD = 1.19$) conditions began with equivalent MSIs, $F_{(1, 44)} = 2.65$, $p = 0.11$. While the post-test scores between the extremely positive ($M = 0.01$, $SD = 0.91$) and extremely negative ($M = 0.10$, $SD = 1.18$) conditions also did not differ by condition, $F_{(1, 44)} = 0.84$, $p = 0.77$, the change between the pre- and post-test did differ by condition, $F_{(1, 44)} = 4.35$, $p = 0.04$, $\eta^2 = 0.09$. Specifically, those who received extremely positive feedback about the states of their moral selves showed an increase between scores on the pre- and post-test ($M = 0.25$, $SD = 0.70$), whereas those who received extremely negative feedback about the states of their moral selves showed a decrease between scores on the pre- and post-test ($M = -0.13$, $SD = 0.48$).

Discussion

As predicted, we found that feedback regarding people's moral selves relative to others led to self-reported changes in their MSI. People who were told they had a moral self that was extremely above average had a positive change between pre- and post-testing, whereas those who were told that they had a moral self that was extremely below average showed a negative change. We wish to acknowledge that while the difference between the two conditions for the pre-test scores was not significant, the extremely positive condition did start at a lower point than the extremely negative condition. This lower pre-test score increased the chances that a mere regression to the mean would produce MSI change scores that would increase for the former condition more so than for the latter. In order to explore the robustness of this effect more thoroughly, in the following two studies, we examine the effects of two additional sources of self-image impact on people's MSI.

STUDY 4

In Study 4, we explore the effect of the second of the three sources of impact to one's self-concept (Kernis and Goldman, 2003), feedback, on changes to a person's MSI (Kernis and Johnson, 1990). Specifically, we examine how explicit feedback about the state of one's moral self relative to one's own personal ideal influences the MSI in both positive and negative ways. To continue the investigation of discriminant validity, we also examined the change in MSI relative to the change in other potentially-related constructs. Specifically, consistent with our argument that feedback about the moral self will only lead to changes to the MSI, we also asked people to assess themselves on four amoral traits (i.e., *sporty, organized, smart, and sociable*). To rule out the possibility that our moral feedback changed people's general self-concept (rather than specifically their MSI), we also examined how our feedback changed people's generalized (Rosenberg, 1965) and state self-esteem (Heatherton and Polivy, 1991). To investigate whether such feedback affected other dimensions of the moral self, we also examined changes to their moral identity (Aquino and Reed, 2002). We predicted that changes following this feedback would only occur on one's MSI and not on the amoral control traits, self-esteem, or moral identity.

Participants

Participants were 130 international business students (52% female, $M_{age} = 21.06$, $SD = 3.03$) at a university in The Netherlands who participated in exchange for €4. We presented all materials in English. Fifteen participants were excluded from the analyses[8], leaving us with a working total of 115 participants.

Design and Procedures

We had three feedback conditions: *meeting ideal moral self*, *almost meeting ideal moral self,* and *a ways away from meeting the ideal moral self*. We chose these types of feedback because they represented people's achievement of their ideal moral self in addition to being both close and far from this ideal state.

[7]Due to a programming error in which the pre-test was measured on a seven-point scale and the post-test was measured on a nine-point scale, all MSI scores were standardized prior to analyses.

[8]Seven participants were excluded because they only took the post-test, one because he/she only took the pre-test, one who took the pre-test after the post-test, one who took the pre-test multiple times, one who's pre-test to post-test difference score was 8 standard deviations above the mean, and four people due to worrisome behavior in the lab (e.g., could not understand the consent form in English, were caught talking on their cell phones in the lab cubicle).

TABLE 5 | Study 4—Pre- and Post-test scale intercorrelations and reliabilities.

	MSI_1	MIs_1	MIi_1	GSE_1	SSE_1	MSI_2	MIs_2	MIi_2	GSE_2	SSE_2	Age	Gender
MSI_1	0.78											
MIs_1	0.29***	0.75										
MIi_1	0.14	0.47***	0.75									
GSE_1	−0.03	−0.07	0.004	0.85								
SSE_1	−0.02	−0.02	−0.08	0.70***	0.86							
MSI_2	0.83***	0.24**	0.06	0.08	0.04	0.89						
MIs_2	0.27**	0.80***	0.42***	−0.02	−0.05	0.22*	0.82					
MIi_2	0.11	0.44***	0.75***	−0.07	−0.16	0.03	0.50***	0.82				
GSE_2	−0.02	−0.02	−0.06	0.86***	0.71***	0.02	0.04	−0.01	0.87			
SSE_2	−0.06	−0.04	−0.06	0.68***	0.89***	−0.01	−0.03	−0.11	0.75***	0.88		
Age	0.19*	0.05	−0.02	−0.02	0.02	0.20*	−0.02	0.02	−0.11	0.00	–	
Gender	−0.17†	−0.13	−0.15	0.28***	0.17†	−0.13	−0.10	−0.13	0.27**	0.18*	0.04	–

*Cronbach alphas contained in the diagonals. †$p < 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; SSE, state self-esteem. $_1$indicates that it was taken in the pre-test. $_2$indicates that it was taken in the post-test. For gender, 1, female; 2, male.*

Participants were required to sign up for the study at least 20 h ahead of their scheduled session. Immediately upon signing up, we sent them a link to an online data collection site, where they completed the pre-test measures: MSI, control trait ratings, moral identity, generalized self-esteem, and state self-esteem. They had to complete these measures at least 15 h in advance of their session in order to participate in the laboratory portion of the study.

Participants came into the lab at least 15 h after completing the pre-test, ostensibly for a study on "e-tests." They were first asked a series of questions about their prosocial behavior and asked to write an essay about what they do to help other people in their daily lives. As in Study 2, we then told them that what they actually just took was a measure called the "MIP," which along with the questions they answered online prior to coming into the lab, indicates how close they are to meeting "the moral self they ideally wish to be." We then gave them both verbal and graphic feedback about where they fell on this scale. Specifically, participants were told that they *met the moral self that they aspire to be*, *have almost met the moral self that they aspire to be*, or were *a ways away from meeting the moral self they aspire to be*, depending on their randomly-assigned condition. We accompanied this feedback with diagrams to show how close they were to their ideal moral self.

Participants then took all the pre-test measures a second time and then were fully debriefed, which included being told that the measure and associated feedback was completely bogus.

Results

All pre- and post-test correlations are contained in **Table 5**. The correlations for the pre-test replicated those found in Study 1 (Samples 1a and 1b) and Study 2 for both the symbolic (positive) and internalized (none) moral identity scores. They also replicated the results found for Study 1 (Sample 1a) for both gender (positive) and age (marginally negative). However, surprisingly, the pre-test MSI was not correlated with generalized self-esteem, as found in the previous studies.

As can be seen in **Table 6**, the MSI changed in the predicted directions based on the feedback we provided, with the *met* feedback raising people's MSI between pre- and post-test and the *a ways away* feedback lowering people's MSI, $F_{(2,\ 112)} = 3.33$, $p = 0.04$, $\eta^2 = 0.06$. However, this was not the case for symbolic moral identity, $F_{(2,\ 112)} = 2.20$, $p = 0.12$, generalized self-esteem, $F_{(2,\ 112)} = 0.01$, $p = 0.99$, state self-esteem, $F_{(2,\ 112)} = 1.41$, $p = 0.25$, and the amoral traits, $F_{(2,\ 112)} = 2.19$, $p = 0.12$. And counter to our predictions, our feedback affected internalized moral identity, $F_{(2,\ 112)} = 3.52$, $p = 0.03$, $\eta^2 = 0.06$, with those receiving the *almost met* feedback showing an increase from pre- to post-test and the *met* condition showing a decrease. There was no significant difference between either of these two conditions and the *a ways away* condition.

Discussion

As we predicted, telling people that they had achieved their ideal moral selves led them to increase their MSI, whereas telling them that they were *a ways away* from achieving their ideal moral selves led them to decrease their MSI. This feedback did not affect people's ratings on the amoral traits, their general or state self-esteem (which is in contrast to previous results, see Barkan et al., 2012), nor their symbolic moral identity.

However, it did affect their internalized moral identity, an unexpected finding both because internalized moral identity is argued to be a stable trait (Aquino and Reed, 2002) and because it has been found to be so in other research (e.g., Jordan et al., 2011). We therefore did not predict that our feedback would change ratings on this construct, which represents the importance that people place on possessing moral traits. Participants placed more importance on possessing moral traits when we told them that they had almost reached their ideal moral selves than when we told them that they had met their ideal moral selves. These results could have been due to an aspiration-level phenomenon (Zhang et al., 2007). In other words, people may have lowered the importance of moral traits when they believed they had met the goal and may have raised the importance when they were told

TABLE 6 | Study 4—Pre- and Post-test scale means and change scores.

Measure	Pre-test mean (SD)	Post-test mean (SD)	*Met* Condition mean (SD)	*Almost Met* Condition mean (SD)	*A ways away* Condition mean (SD)
MSI	5.06 (0.80)	5.01 (0.90)	0.11_{a} (0.46)	$-0.07_{a,b}$ (0.51)	-0.18_{b} (0.51)
MIs	5.74 (0.83)	5.77 (0.83)	−0.20 (0.60)	−0.10 (0.66)	0.09 (0.64)
MIi	4.14 (1.02)	4.07 (1.03)	-0.14_{a} (0.59)	0.21_{b} (0.66)	$0.02_{a,b}$ (0.47)
GSE	3.78 (0.57)	3.76 (0.56)	−0.02 (0.26)	−0.01 (0.27)	−0.01 (0.37)
SSE	3.40 (0.49)	3.47 (0.50)	0.03 (0.26)	0.12 (0.22)	0.05 (0.22)
Control traits	4.52 (0.98)	4.53 (0.90)	0.04 (0.57)	−0.12 (0.51)	0.12 (0.43)

MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; SSE, state self-esteem. For those variables with a significant omnibus ANOVA, means with different subscripts significantly differ at a $p < 0.05$.

that they were "almost there." Future, research on the variance of internalized moral identity should investigate this possibility.

As discussed earlier, in addition to feedback, people's self-concepts are influenced by their own actions (Kernis and Goldman, 2003). As such, people's MSI should also respond to their moral actions and to their recalls about their moral actions (Sachdeva et al., 2009; Jordan et al., 2011). Thus, the purpose of Study 5 is to examine how recall of moral and immoral behavior changes people's MSI.

STUDY 5

Study 5 aimed to explore the third source of influence on one's self concept—an individual's actions (Kernis and Goldman, 2003)—by analyzing whether the recall of one's (im)moral actions alters one's MSI. Cornelissen et al. (2013) used the current MSI scale to demonstrate that the effects of recalling one's (im)moral actions on future immoral behavior can be explained by the state of one's MSI. Specifically, they asked people to recall a time when they acted in a way that intentionally harmed another person (immoral) or intentionally benefitted another person (moral). They then had them engage in a task where they could cheat for their own personal gain (adapted from Mazar et al., 2008) and found that people who recalled harming another person cheated on fewer tasks than those who recalled helping another person and that these compensatory effects were explained by the level of a person's MSI, as measured by the current scale[9]. This finding would suggest that the change to one's MSI caused by (im)moral actions explains people's subsequent moral compensation behavior.

However, and as noted before, although these authors used the current scale to demonstrate this mediation, they did not demonstrate that people's MSI actually changed from a baseline; they also did not compare that effect to other possible changes in similarly related constructs[10]. Thus, in Study 4, we used these exact recalls to examine how they changed people's MSI from a baseline level. We also included a control condition to examine the directionality of the effects. We predicted that whereas recalling an immoral action would lower people's MSI, recalling a moral action would raise people's MSI.

Participants

Participants were 119 international business students (48% female, $M_{age} = 21.68$, $SD = 2.96$) at a university in The Netherlands who participated in exchange for €4. All materials were presented in English. We excluded 12 people, leaving us with a working total of 107 participants[11].

Design and Procedures

We had two conditions that were identical to those used by Cornelissen et al. (2013): recalling an intentional action one engaged in that harmed another person or recalling an intentional action one engaged in that benefitted another person. For example, in the unethical condition, participants wrote about behaviors such as borrowing money from another person and then waiting until the other person likely had forgotten so that he/she did not have to pay the person back, or delivering low-quality work on a group project in the expectation that other members would compensate for it. In the ethical condition, participants wrote about behaviors such as loaning a friend money that one had set aside for new clothes or joining a friend for an event that the other person did not feel comfortable attending alone despite being tired. We also included a control condition in which participants were asked to recall their last visit to the grocery store.

All other procedures were identical to those used in Study 4: participants were required to complete the pre-test measures (i.e., MSI, control traits, moral identity, state self-esteem, and generalized self-esteem) at least 15 h in advance of their session in order to participate in the laboratory portion of the study. In the laboratory session, they completed the recall task and then all pre-test measures once again. Finally, they were fully debriefed.

Results

All pre- and post-test correlations are included in **Table 7**. The correlations for the pre-test replicated those found in S1 (Samples

[9]They also had people recall times when they violated or acted consistently with a moral rule. These recalls did not lead to compensatory effects on immoral behavior and the relationship was not mediated by the MSI.

[10]Note that the scale in the Cornelissen et al. (2013) paper was not created by these authors themselves, but was the MSI scale presented in the current manuscript (cited from a conference presentation of this scale; see Footnote 2).

[11]Four participants were excluded because they did not take the pre-test, one because he/she did not take the post-test, six people who took the pre-test significantly less than 15 h before the post-test, and one person due to worrisome behavior in the lab (i.e., read the study debriefing before going in for the post-test).

TABLE 7 | Study 5—Pre- and Post-test scale intercorrelations and reliabilities.

	MSI_1	MIs_1	MIi_1	GSE_1	SSE_1	MSI_2	MIs_2	MIi_2	GSE_2	SSE_2	Age	Gender
MSI_1	0.72											
MIs_1	0.32***	0.64										
MIi_1	0.12	0.46***	0.72									
GSE_1	0.09	0.04	−0.04	0.80								
SSE_1	−0.04	−0.02	−0.12	0.73***	0.86							
MSI_2	0.75***	0.35***	0.25**	0.02	−0.03	0.79						
MIs_2	0.21**	0.78***	0.49***	−0.06	−0.09	0.29**	0.76					
MIi_2	0.15	0.43***	0.78***	−0.07	−0.14	0.24*	0.48***	0.78				
GSE_2	0.00	−0.02	−0.02	0.86***	0.73***	0.04	−0.07	0.03	0.78			
SSE_2	−0.01	−0.03	−0.03	0.68***	0.84***	−0.02	−0.07	−0.06	0.70***	0.84		
Age	0.01	−0.01	−0.02	−0.09	−0.07	0.02	0.08	0.08	−0.10	−0.09	–	
Gender	−0.02	0.04	−0.10	0.05	0.08	0.02	−0.02	−0.06	0.15	0.05	−0.01	–

*Cronbach alphas contained in the diagonals. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$. MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; SSE, state self-esteem. $_1$indicates that it was taken in the pre-test. $_2$indicates that it was taken in the post-test. For gender, 1, female; 2, male.*

TABLE 8 | Study 5—Pre- and Post-test scale means and change scores.

Measure	Pre-test mean (SD)	Post-test mean (SD)	*Moral* Condition mean (SD)	*Immoral* Condition mean (SD)	Control condition mean (SD)
MSI	5.06 (0.73)	5.05 (0.76)	0.11_a (0.40)	-0.21_b (0.59)	$-0.02_{a,b}$ (0.55)
MIs	4.13 (0.94)	3.93 (0.90)	−0.20 (0.61)	−0.26 (0.60)	−0.14 (0.63)
MIi	5.71 (0.72)	5.50 (0.84)	−0.27 (0.60)	−0.17 (0.38)	−0.16 (0.59)
GSE	3.86 (0.50)	3.87 (0.46)	−0.02 (0.26)	0.03 (0.24)	0.02 (0.29)
SSE	3.47 (0.47)	3.55 (0.43)	0.06 (0.26)	0.10 (0.26)	0.04 (0.25)
Control Traits	4.49 (0.97)	4.60 (0.91)	0.24 (0.51)	−0.06 (0.59)	0.14 (0.83)

MSI, moral self-image; MIs, moral identity–symbolization; MIi, moral identity–internalization; GSE, generalized self-esteem; SSE, state self-esteem. For those variables with a significant omnibus ANOVA, means with different subscripts significantly differ at a $p < 0.05$.

1a and 1b), Study 2, and Study 4 for both the symbolic (positive) and internalized (none) moral identity scores. They, however, failed to show any effects for either gender or age. And as found in Study 4 (albeit not Study 1 or 2), the pre-test MSI was not correlated with generalized self-esteem. We discuss possible reasons for this in the General Discussion.

As can be seen in **Table 8**, the MSI changed in the predicted directions based on people's recalled situations, $F_{(2,\ 103)} = 3.79$, $p = 0.03$, $\eta^2 = 0.07$: recalls of people's moral actions led to increases in their MSI, recalls of people's immoral actions led to decreases in their MSI, and the control recall led to virtually no change. However, this was not the case for the moral identity measure [both symbolic, $F_{(2,\ 103)} = 0.20$, $p = 0.70$, and internalized, $F_{(2,\ 103)} = 0.46$, $p = 0.64$], generalized self-esteem, $F_{(2,\ 103)} = 0.31$, $p = 0.73$, state self-esteem, $F_{(2,\ 103)} = 0.48$, $p = 0.62$, and the control traits, $F_{(2,\ 103)} = 2.24$, $p = 0.11$.

Discussion

The goal of Study 5 was to determine if the prompts used by Cornelissen et al. (2013) that altered people's immoral behavior actually changed their MSI. We indeed found that the recall of one's (im)moral behavior changed one's MSI in the predicted directions. Also as predicted, this recall did not affect people's assessment on the control traits or their moral identities—including their symbolic moral identity, internalized moral identity (which is inconsistent with Study 4 but consistent with initial predictions), state self-esteem (which, again, is inconsistent with what previous research has found, see Barkan et al., 2012), and generalized self-esteem.

These findings appear to be consistent with the theory of moral compensation as symbolic self-completion (Zhong and Liljenquist, 2006; Jordan et al., 2011). That is, moral actions (or recalled moral actions) raise people's MSI, thus allowing them to relax their strivings on subsequent moral tasks. Similarly, immoral actions (or recalled immoral actions) lower people's MSI, thus leading them to put greater effort into acting morally on subsequent tasks (Sachdeva et al., 2009).

GENERAL DISCUSSION

While we have witnessed people's moral inconsistencies both in real life and experimental research (e.g., Monin and Miller, 2001; Zhong and Liljenquist, 2006; Sachdeva et al., 2009; Jordan et al., 2011), until now, there was no validated measure to empirically examine the impact of these inconsistencies on people's MSI nor to examine the potential psychological processes driving these inconsistencies. As we propose in the current manuscript, these moral behaviors impact people's MSI in positive and negative ways. And as others have demonstrated (e.g., Cornelissen et al., 2013), these effects on the MSI subsequently affect related moral behaviors; in other words, MSI is a malleable construct that helps explain (im)moral behavior, like generosity and dishonesty.

This investigation accomplished two important objectives. First, it developed a scale to measure the MSI and, in order to investigate its convergent and discriminant validity, conceptually and empirically compared it to related constructs, such as moral identity (Aquino and Reed, 2002) and self-esteem (Rosenberg, 1965). Despite its theoretical relationship to both moral identity and self-esteem, we found that the MSI was empirically distinct from these constructs (Studies 1 and 2). It should be acknowledged that we did not find a relationship between the MSI and generalized self-esteem in either Study 4 or 5, which is curious since we found a relationship in the previous three studies in which generalized self-esteem was administered. A potential reason for this may be the samples used. The studies in which we found relationships between the MSI and generalized self-esteem employed non-student American adult samples, whereas those that did not, used Dutch student samples. It is possible that that given the secularism of Western Europe (Berger et al., 2008), Dutch students did not feel a connection between the state of their MSI and their general self-image. It could also be an age-related effect, such that in early adulthood, one's perceived moral state feels fairly isolated from his or her general self-image. In order to understand these effects further, more in-depth exploration of this issue is needed. Second, and relatedly, while the MSI was affected by three sources of influence (Kernis and Goldman, 2003)—social comparison, explicit feedback, and one's own behavior—this feedback did not affect these other constructs (with the exception of internalized moral identity in Study 4).

The current research also has implications outside the laboratory. Specifically, it suggests that specific events and feedback from the environment can affect people's MSI. This means that events in in the social world, such as reflecting on one's moral or immoral behavior during an interaction, can affect how an individual perceives his or her moral self. It also means that feedback about one's moral or immoral behavior, which routinely comes from experiences such as organizational, school, or family life, can affect the state of one's MSI.

Limitations and Future Directions

There are limitations of the current investigation that warrant acknowledgment. First, in Study 4 we found an effect of feedback on internalized moral identity; feedback that one had almost reached one's ideal moral self increased one's reported importance of possessing moral traits (i.e., internalized moral identity), whereas feedback that one had met his or her ideal moral self led the individual to decrease such reported importance. This finding was unexpected given that prior research found internalized moral identity to be a stable trait (Jordan et al., 2011), and it is conceptualized as such (Aquino and Reed, 2002; Aquino et al., 2009). It is possible that the aspirational-level phenomenon (Zhang et al., 2007) may explain this result; however, more research is needed to investigate this and other possible explanations, as we did not find this effect in Study 5. Relatedly, in Studies 3 through 5, in which we either manipulated feedback about people's moral selves or allowed people to reflect on their own moral behavior, we always placed the MSI scale before the other scales, as observing changes to the MSI constituted the main goal of these studies. Therefore, we wished to minimize and distractions for participants between the presentation of the manipulations and people's ratings of their moral selves. We acknowledge that this methodology may have biased the results in favor of finding changes to people's MSI rather than to the scales that came later in the line-up (e.g., moral identity or state self-esteem).

Second, it is possible that the traits we used to capture the MSI were not traits that universally corresponded to people's conceptualization of the ideal moral person. For example, there is evidence that the connection between work and morality is specific to cultures with puritanical, Calvinist origins (Uhlmann et al., 2011). Thus, the trait, *hardworking,* might not elicit a prototype of the moral person equally across cultures. Therefore, it is possible that not all people collectively viewed these nine traits as equivalently referential to the moral self. While we used diverse samples to demonstrate our results, from American adults to international students, future research is needed to understand cultural differences on the conceptualization of moral prototypes.

Third, although the MSI is a state scale, we did not instruct people to rate how they were feeling "right now"—that is, at the current moment. Thus, it is possible, that some people rated themselves on these traits based on how they felt about their MSI, in general. However, results from Studies 3 through 5 did demonstrate variance between pre- and post-tests of individuals' MSI, suggesting that they were rating themselves based on perceptions at the current moment. However, it also suggests that leaving this phrasing out of the scale's preamble meant that our results served as a conservative test of our theory and that bigger pre- to post-test discrepancies may have been found had we emphasized the construct's state nature in our phrasing. We encourage future researchers using the MSI scale to experiment with the use of the "right now" statement and explore how it affects participants' responses.

An additional future direction would be to investigate the interaction between MSI and moral identity. There is suggestive evidence that MSI might interact with moral identity to affect people's engagement in moral behavior. For example, it might be that only when internalized moral identity is high (that is, when a person highly values possessing moral traits) does a low MSI prompt moral behavior in order to restore the moral self. As Aquino et al. (2009) wrote, "someone whose self-definition is organized around a set of moral traits should be motivated to behave in a moral manner to maintain this self-conception" (p. 124). They also hypothesized that people with a lower internalized moral identity would not be prompted to show such restorative behaviors. That said, there may be some empirical difficulties in testing this hypothesis due to ceiling effects, as the mean internalized moral identity is consistently found to be quite high (e.g., a 4.6 on a five-point scale, Aquino and Reed, 2002, and a 6.28 on a seven-point scale, Reed et al., 2007).

CONCLUSION

Thinking about countless societal examples, a person can be both a pillar of the community and a thief, engaging in reflections that likely both boost and lower the way she thinks about her moral self. The current investigation demonstrates that the MSI

is malleable and also presents a way to gauge this malleability with the goal of providing researchers with a more nuanced understanding of the intersection between the moral self and moral behavior.

FUNDING

This research was supported by a grant (#451-13-031) awarded to the second author by The Netherlands Organization for Scientific Research (NWO).

ACKNOWLEDGMENTS

We would like to thank Francesca Gino for her contribution to this project and George Newman for his comments on an earlier draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01878

REFERENCES

Adler, A. (2006). "Fundamentals of individual psychology," in *Readings in the Theory of Individual Psychology*, eds S. Slavik and J. Carlson (New York, NY: Routledge/Taylor & Francis Group), 33–43.

Ahmed, S. A., and Jackson, D. N. (1979). Psychographics for social policy decisions: welfare assistance. *J. Consumer Res.* 5, 229–239. doi: 10.1086/208735

Allport, G. W. (1955). *Becoming; Basic Considerations for a Psychology of Personality*. New Haven, CT: Yale University Press.

Aquino, K., Freeman, D., Reed, I. I., Americus, Lim, V. K. G., and Felps, W. (2009). Testing a social-cognitive model of moral behavior: the interactive influence of situations and moral identity centrality. *J. Pers. Soc. Psychol.* 97, 123–141. doi: 10.1037/a0015406

Aquino, K., and Reed, A. (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83, 1423–1440. doi: 10.1037/0022-3514.83.6.1423

Aronson, E. (2003). *The Social Animal*. New York, NY: Macmillan.

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *J. Exp. Psychol. Gen.* 141, 757–773. doi: 10.1037/a0027588

Bazerman, M. H., and Gino, F. (2012). Behavioral ethics: toward a deeper understanding of moral judgment and dishonesty. *Ann. Rev. Law Soc. Sci.* 8, 85–104. doi: 10.1146/annurev-lawsocsci-102811-173815

Berger, P. L., Davie, G., and Fokas, E. (2008). *Religious America, Secular Europe?: A Theme and Variation*. Hampshire, UK: Ashgate Publishing, Ltd.

Brown, J. D. (1993). "Self-esteem and self-evaluation: feeling is believing," in *Psychological Perspectives on the Self*, Vol. 4, ed J. Suls (Hillsdale, NJ: Erlbaum), 27–58.

Brown, L. B. (1962). A study of religious belief. *Br. J. Psychol.* 53, 259–272. doi: 10.1111/j.2044-8295.1962.tb00832.x

Colby, A., Kohlberg, L., Gibbs, J., Lieberman, M., Fischer, K., and Saltzstein, H. D. (1983). A longitudinal study of moral judgment. *Monogr. Soc. Res. Child Dev.* 48, 1–124. doi: 10.2307/1165935

Cornelissen, G., Bashshur, M. R., Rode, J., and Menestrel, M. L. (2013). Rules or consequences? the role of ethical mind-sets in moral dynamics. *Psychol. Sci.* 24, 482–488. doi: 10.1177/0956797612457376

Crocker, J., and Wolfe, C. T. (2001). Contingencies of self-worth. *Psychol. Rev.* 108:593. doi: 10.1037/0033-295X.108.3.593

Damon, W., and Hart, D. (1992). "Self-understanding and its role in social and moral development," in *Developmental Psychology: An Advanced Textbook, 3rd Edn.*, eds M. H. Bornstein and M. E. Lamb (Hillsdale, NJ: Lawrence Erlbaum Associates), 421–464.

Deci, E. L., and Ryan, R. M. (1995). "Human autonomy: the basis for true self-Esteem," in *Efficacy, Agency, and Self-Esteem*, ed M. H. Kernis (New York, NY: Plenum Press), 31–49.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70:979. doi: 10.1037/0022-3514.70.5.979

Dunning, D. (2007). Self-image motives and consumer behavior: how sacrosanct self-beliefs sway preferences in the market place. *J. Consum. Psychol.* 17, 237–249. doi: 10.1016/S1057-7408(07)70033-5

Edwards, J. R. (2002). "Alternatives to difference scores: polynomial regression analysis and response surface methodology," in *Advances in Measurement and Data Analysis*, eds F. Drasgow and N. W. Schmitt (San Francisco, CA: Jossey-Bass), 350–400.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annu. Rev. Psychol.* 51, 665–697. doi: 10.1146/annurev.psych.51.1.665

Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., and Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *J. Res. Adolesc.* 15, 235–260. doi: 10.1111/j.1532-7795.2005.00095.x

Eisenberger, R., Lynch, P., Aselage, J., and Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Pers. Soc. Psychol. Bull.* 30, 787–799. doi: 10.1177/0146167204264047

Fiske, S. T., Cuddy, A. J., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Fiske, S. T., Xu, J., Cuddy, A. C., and Glick, P. (1999). (Dis)respecting versus (Dis)liking: status and interdependence predict ambivalent stereotypes of competence and warmth. *J. Soc. Issues* 55, 473–489. doi: 10.1111/0022-4537.00128

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Goldstone, R. L., and Chin, C. (1993). Dishonesty in self-report of copes made: moral relativity and the copy machine. *Basic Appl. Soc. Psych.* 14, 19–32. doi: 10.1207/s15324834basp1401_2

Gouldner, A. (1960). The norm of reciprocity: a preliminary statement. *Am. Soc. Rev.* 25, 161–178. doi: 10.2307/2092623

Greenwald, A. G., and Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *J. Pers. Soc. Psychol.* 79:1022. doi: 10.1037/0022-3514.79.6.1022

Heatherton, T. F., and Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *J. Pers. Soc. Psychol.* 60, 895–910. doi: 10.1037/0022-3514.60.6.895

Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychol. Rev.* 94, 319–340. doi: 10.1037/0033-295X.94.3.319

Hoffman, M. L. (1975). Sex differences in moral internalization and values. *J. Pers. Soc. Psychol.* 32:720. doi: 10.1037/0022-3514.32.4.720

Jaffee, S., and Hyde, J. S. (2000). Gender differences in moral orientation: a meta-analysis. *Psychol. Bull.* 126, 703–726. doi: 10.1037/0033-2909.126.5.703

Jones, S. C. (1973). Self-and interpersonal evaluations: esteem theories versus consistency theories. *Psychol. Bull.* 79, 185. doi: 10.1037/h0033957

Jordan, J., Mullen, E., and Murnighan, J. K. (2011). Striving for the moral self: the effects of recalling past moral actions on future moral behavior. *Personal. Soc. Psychol. Bull.* 37, 701–713. doi: 10.1177/0146167211400208

Kernis, M. H., Cornell, D. P., Sun, C. R., Berry, A., and Harlow, T. (1993). There's more to self-esteem than whether it is high or low: the importance of stability of self-esteem. *J. Person. Soc. Psychol.* 65:1190. doi: 10.1037/0022-3514.65.6.1190

Kernis, M. H., and Goldman, B. M. (2003). "Stability and variability in self-concept and self-esteem," in *Handbook of Self and Identity,* eds M. R. Leary and J. P. Tangney (New York, NY: Guilford Press), 106–127.

Kernis, M. H., and Johnson, E. K. (1990). Current and typical self-appraisals: differential responsiveness to evaluative feedback and implications for emotions. *J. Res. Pers.* 24, 241–257. doi: 10.1016/0092-6566(90)90019-3

Kohlberg, L. (1971). Stages of moral development. *Moral Educ.* 23–92.

Kohlberg, L. (1994). "Stage and sequence: the cognitive-developmental approach to socialization," in *Defining Perspectives in Moral Development,* ed B. Puka (New York, NY: Garland Publishing), 1–134.

Kohlberg, L., Levine, C., and Hewer, A. (1983). *Moral Stages: A Current Formulation and a Response to Critics.* New York, NY; Karger.

Kotter, J. P. (1973). The psychological contract: managing the joining-up process. *Calif. Manage. Rev.* 15, 91–99. doi: 10.2307/41164442

Lapsley, D. K., and Lasky, B. (2001). Prototypic moral character. *Identity An Int. J. Theory Res.* 1, 345–363. doi: 10.1207/S1532706XID0104_03

Lewicki, P. (1983). Self-image bias in person perception. *J. Pers. Soc. Psychol.* 45:384. doi: 10.1037/0022-3514.45.2.384

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Market. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mazar, N., and Zhong, C. B. (2010). Do green products make us better people? *Psychol. Sci.* 21, 494–498. doi: 10.1177/0956797610363538

McCarthy, J. D., and Hoge, D. R. (1982). Analysis of age effects in longitudinal studies of adolescent self-esteem. *Dev. Psychol.* 18:372. doi: 10.1037/0012-1649.18.3.372

Monin, B., and Jordan, A. H. (2009). "The dynamic moral self: a social psychological perspective," in *Personality, Identity, and Character: Explorations in Moral Psychology,* eds D. Narvaez and D. K. Lapsley (New York, NY: Cambridge University Press), 341–354.

Monin, B., and Miller, D. T. (2001). Moral credentials and the expression of prejudice. *J. Pers. Soc. Psychol.* 81, 33–43. doi: 10.1037/0022-3514.81.1.33

Moore, C. (2008). Moral disengagement in processes of organizational corruption. *J. Bus. Ethics* 80, 129–139. doi: 10.1007/s10551-007-9447-8

Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., and Mayer, D. M. (2012). Why employees do bad things: moral disengagement and unethical organizational behavior. *Pers. Psychol.* 65, 1–48. doi: 10.1111/j.1744-6570.2011.01237.x

Moore, C., and Gino, F. (2013). Ethically adrift: how others pull our moral compass from true north, and how we can fix it. *Res. Organ. Behav.* 33, 53–77. doi: 10.1016/j.riob.2013.08.001

Moretti, M. M., and Higgins, E. T. (1990). Relating self-discrepancy to self-esteem: the contribution of discrepancy beyond acutal-self ratings. *J. Exp. Soc. Psychol.* 26, 108–123. doi: 10.1016/0022-1031(90)90071-S

Mulder, L. B., and Aquino, K. (2013). The role of moral identity in the aftermath of dishonesty. *Organ. Behav. Hum. Decis. Process.* 121, 219–230. doi: 10.1016/j.obhdp.2013.03.005

Murnighan, J. K., Kim, J. W., and Metzger, A. R. (1993). The volunteer dilemma. *Adm. Sci. Q.* 38, 515–538. doi: 10.2307/2393335

Pelham, B. W. (1995). Self-investment and self-esteem: evidence for a Jamesian model of self-worth. *J. Pers. Soc. Psychol.* 69:1141. doi: 10.1037/0022-3514.69.6.1141

Reed, A., Aquino, K., and Levy, E. (2007). Moral identity and judgments of charitable behaviors. *J. Mark.* 71, 178–193. doi: 10.1509/jmkg.71.1.178

Rosenberg, M. (1965). *Society and the Adolescent Self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiveing the Self*. New York, NY: Basic Books.

Rosenberg, M. (1986). "Self-concept from middle childhood through adolescence," in *Psychological Perspectives on the Self*, Vol. 2, eds J. Suls and G. Greenwald (Hillsdale, NJ: Erlbaum), 107–136.

Sachdeva, S., Iliev, R., and Medin, D. L. (2009). Sinning saints and saintly sinners: the paradox of moral self-regulation. *Psychol. Sci.* 20, 523–528. doi: 10.1111/j.1467-9280.2009.02326.x

Schweitzer, M. E., Ordonez, L., and Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Acad. Manag. J.* 47, 422–432. doi: 10.2307/20159591

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Steele, C. M. (1988). "The psychology of self-affirmation: sustaining the integrity of the self," in *Advances in Experimental Social Psychology, Vol. 21, Social Psychological Studies of the Self: Perspectives and Programs,* ed L. Berkowitz (San Diego, CA: Academic Press), 261–302.

Uhlmann, E. L., Poehlman, T. A., Tannenbaum, D., and Bargh, J. A. (2011). Implicit puritanism in american moral cognition. *J. Exp. Soc. Psychol.* 47, 312–320. doi: 10.1016/j.jesp.2010.10.013

Wark, G. R., and Krebs, D. L. (1996). Gender and dilemma differences in real-life moral judgment. *Dev. Psychol.* 32, 220–230. doi: 10.1037/0012-1649.32.2.220

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

White, R. D. (1999). Are women more ethical? recent findings on the effects of gender upon moral development. *J. Public Admin. Res. Theory* 9, 459–472. doi: 10.1093/oxfordjournals.jpart.a024418

Wicklund, R. A., and Gollwitzer, P. M. (1981). Symbolic self-completion, attempted influence, and self-deprecation. *Basic Appl. Soc. Psych.* 2, 89–114. doi: 10.1207/s15324834basp0202_2

Wylie, R. C. (1974). *The Self-Concept: Theory and Research on Selected Topics, Vol. 2*. Lincoln, NB: University of Nebraska Press.

Zhang, Y., Fishbach, A., and Kruglanski, A. W. (2007). The dilution model: how additional goals undermine the perceived instrumentality of a shared path. *J. Pers. Soc. Psychol.* 92:389. doi: 10.1037/0022-3514.92.3.389

Zhong, C.-B., and Liljenquist, K. (2006). Washing away your sins: threatened morality and physical cleansing. *Science* 313, 1451–1452. doi: 10.1126/science.1130726

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 July 2016
doi: 10.3389/fpsyg.2016.01063

The Effect of Self-Esteem on Corrupt Intention: The Mediating Role of Materialism

Yuan Liang, Li Liu, Xuyun Tan, Zhenwei Huang, Jianning Dang and Wenwen Zheng*

Beijing Key Lab of Applied Experimental Psychology, School of Psychology, Beijing Normal University, Beijing, China

The present set of studies aimed to explore the effect of self-esteem on corrupt intention and the mediating role of materialism in generating this effect. In Study 1, we used questionnaires to investigate the correlation among self-esteem, materialism, and corrupt intention. In Study 2, we manipulated self-esteem to explore the causal effect of self-esteem on materialism and corrupt intention. In Study 3, we manipulated materialism to examine whether inducing materialism can reduce the relationship between self-esteem and corrupt intention. The three studies converged to show that increased self-esteem caused a low level of materialism, which in turn decreased corrupt intention. The theoretical and practical implications of the results are discussed.

Keywords: self-esteem, corrupt intention, materialism, unethical behavior, psychological process

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Amos Schurr,
Ben-Gurion University of the Negev, Israel
Andrea Pittarello,
Ben-Gurion University of the Negev, Israel

****Correspondence:***
Li Liu
l.liu@bnu.edu.cn

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *27 December 2015*
Accepted: *29 June 2016*
Published: *12 July 2016*

Citation:
Liang Y, Liu L, Tan X, Huang Z, Dang J and Zheng W (2016) The Effect of Self-Esteem on Corrupt Intention: The Mediating Role of Materialism. Front. Psychol. 7:1063. doi: 10.3389/fpsyg.2016.01063

INTRODUCTION

Corruption refers to the misuse of public power for private gain (Treisman, 2000). It has brought great negative influences on our society. For instance, corruption increases the cost of doing business by up to 10% on average, and decreases 5% investment in corruptive countries than that in relatively corruption-free countries. It is estimated that the annual cost of corruption worldwide is US$2.6 trillion with over US$1 trillion paid in bribes each year (Organization for Economic Co-operation and Development [OECD], 2014). In addition, citizens turning out at elections in very corrupt countries are 20–30% fewer compared with countries with little corruption (Stockemer et al., 2013). Given that such heavy cost is hard to afford, it is important to figure out the psychological underpinnings of corruption, and thereby to provide scientific support for fighting against corruption.

Corruption has recently received greater attention from psychologists. The subsequent research has identified important antecedents of corruption such as socio-economic status (Olken, 2009; Charron, 2016), cultural orientation (Li et al., 2006; Mazar and Aggarwal, 2011; Huang et al., 2015), risk attitude (Berninghaus et al., 2013), and social dominance orientation (Tan et al., 2016b). Inspired by the ideas that self-esteem could be maintained and enhanced by material possessions or prestige (Sivanathan and Pettit, 2010; Jiang et al., 2015), and that corruption offers a rapid way to obtain admirable status and material resources (You, 2007; Johnston, 2012), the current research investigates an issue that has been overlooked: the effect of self-esteem on corruption. Namely, it aims to address two questions: Does self-esteem influence corrupt intention? If so, what is the underlying psychological process of the association?

Self-Esteem and Corrupt Intention

Self-esteem refers to the overall self-evaluation of one's worth, and it is a universal and fundamental human need (Allport, 1955; Maslow, 1968; Taylor and Brown, 1988; Solomon et al., 1991;

Baumeister et al., 1993). Most individuals aim to protect, maintain, and enhance their self-esteem (Baumeister, 1998). Individuals with high self-esteem believe they can succeed and enhance themselves based on their own merits and are less concerned with avoiding failure (Blaine and Crocker, 1993; Vohs and Heatherton, 2001). By contrast, individuals with low self-esteem feel inferior, unworthy, lonely, insecure, anxious and depressed, uncertain about themselves, and particularly challenged to succeed, and they interpret events and feedback in terms of what they indicate about their self (Mruk, 1995; Brown and Marshall, 2001; Pyszczynski et al., 2004; Donnellan et al., 2005). According to Self-Affirmation Theory (Steele, 1988), when people feel uncertain in one domain, they compensate for this by "spontaneously emphasizing certainty and conviction about unrelated attitudes, values, personal goals, and identifications" (McGregor et al., 2001, p. 473). This compensation constitutes part of a hydraulic motivational process called compensatory conviction. Therefore, individuals whose self-esteem is threatened are motivated to seek any boost to compensate for low self-esteem (Crocker and Park, 2004).

Achieving rewards and status could facilitate self-affirmation and likewise enhance self-esteem for individuals lacking it (Sivanathan and Pettit, 2010). Many unethical behaviors, such as corruption, can be performed to facilitate such achievement. Corruption, like other dishonest acts, is not only motivated by external benefits, but also by internal rewards (Mazar et al., 2008). In order to enhance self-esteem, individuals with low self-esteem divert their attention from fulfilling intrinsic fundamental human needs to pursuing extrinsic outcomes, which pursuit exacerbates already poor self-regulation (Baumeister and Leary, 1995; Deci and Ryan, 2000; Crocker, 2002). Thus, these individuals are more likely to adopt risky, self-aggrandizing, get-rich-quick schemes (Tice, 1993; Zywica and Danowski, 2008) to secure admiration. Corruption offers a rapid way to obtain admirable status and material resources (You, 2007; Johnston, 2012), and even though these items are not theirs (Ledeneva, 1998), they can enhance self-esteem (Richins and Dawson, 1992; Chang and Arkin, 2002; Wattanasuwan, 2005; Isaksen and Roper, 2012). Therefore, individuals with depressed self-esteem prefer to use corruption as a crutch to enhance their self-worth. By contrast, positive self-esteem is not in desire for enhancement, which leads individuals to adhere to ethical standards rather than engage in corruption (Aronson and Mettee, 1968; Tice, 1993; Mesmer-Magnus et al., 2010; Barkan et al., 2015; Jordan et al., 2015). Thus, we hypothesize that *high self-esteem decreases corrupt intention (Hypothesis 1).*

Materialism as a Mediator

If one is increasingly driven by extrinsic goals such as wealth, possessions, image, and status to affirm the self and to seek compensation for poor self-esteem, one might be mired by materialism. Materialism refers to the elevated importance placed by a person on possessions and their acquisition as a necessary or a desirable means of attaining desired end states, including happiness (Richins and Dawson, 1992). Research has indicated that materialism can be used to compensate for threatened self-esteem (Shrum et al., 2013; Jiang et al., 2015) and may prompt unethical behavior (Kouchaki et al., 2013; Gino and Mogilner, 2014). It is thus reasonable to assume that materialism might mediate the effect of self-esteem on corrupt intention. If materialism accounts for the effect of self-esteem on corrupt intention, then we can block materialism to control corruption in individuals with low self-esteem.

Indeed, the earliest sophisticated attempt to measure materialism (Belk, 1985) conceived of the construct as a trait, and more recent theoretical statements have proposed that materialism is an aspect of identity (Dittmar, 2008; Shrum et al., 2013). By contrast, following Richins and Dawson (1992), the current set of studies regards materialism as value or a goal that reflects the extent to which an individual believes acquiring money or possessions is important. It also conveys the striving for the related objects of an appealing image and a high status/popularity, both of which objects are frequently expressed through money and possessions (Kasser, 2016). Understanding materialism as a value/goal allows us to test hypotheses about both a person's relatively stable disposition toward materialism and what occurs when materialistic values/goals are momentarily activated in a person's mind.

It has been shown that self-esteem is negatively associated with materialism (Isaksen and Roper, 2012; Kasser et al., 2014). Self-esteem helps individuals respond to self-worth threats by emphasizing their competence or dominance and become more independent (Blaine and Crocker, 1993; Vohs and Heatherton, 2001). Individuals with high self-esteem do not require many material possessions for purposes of gaining a certain status, image, admiration (Richins and Dawson, 1992), obtaining ephemeral economic safety (Christopher et al., 2006; Clark et al., 2011), restoring psychological security (Noguti and Bokeyar, 2014), affirming one's self-identity (Chang and Arkin, 2002; Wattanasuwan, 2005), or coping with doubts concerning self-worth or competence (Chang and Arkin, 2002; Kasser, 2002; Jiang et al., 2015). Individuals with low self-esteem are in contrast likely to use more money to compensate for their impaired self-esteem (Shrum et al., 2013; Jiang et al., 2015) and to require prestige and many possessions to identify themselves (Richins and Dawson, 1992; Magee and Galinsky, 2008; Mogilner and Aaker, 2009). This necessity thus generates a powerful inner drive to acquire many impressive belongings. This pattern implies that high self-esteem might decrease materialism.

Existing literature indicates that materialism is positively associated with unethical behaviors (Tang and Chiu, 2003; Tang et al., 2008; Tang and Liu, 2011). Simply primed with money leads individuals to be less helpful and less fair when they interact with others, work harder toward their personal goals (Vohs et al., 2006, 2008; Zhou et al., 2009; Yang et al., 2013; Gino and Mogilner, 2014), and even engage in unethical behaviors (Kouchaki et al., 2013). Empirical evidence indicates that individuals with high levels of materialism are more self-oriented; more focused on wealth, achievement, power, and status; and less concerned with others (Richins and Dawson, 1992; Schwartz, 1992; Sheldon and McGregor, 2000; Bauer et al., 2012; Gino and Mogilner, 2014). Materialism has been associated with consumers actively favoring the benefits of illegal actions: a highly materialistic consumer is more likely to tolerate unethical actions if they enhance personal

material possessions or reduce the material possessions of others (Chowdhury and Fernando, 2013). This association implies that materialism might increase corrupt intention.

Based on the above arguments, we hypothesize that *materialism mediates the negative effect of self-esteem on corrupt intention (Hypothesis 2).*

In current research, we conducted three studies with correlated and experimental designs to test whether high self-esteem decreases corrupt intention and whether materialism plays a mediating role in the relationship between self-esteem and corrupt intention. In Study 1, we used questionnaires to investigate the correlated relationship among self-esteem, materialism, and corrupt intention. In Study 2, we manipulated self-esteem to explore its causal effect on materialism and corrupt intention. In Study 3, we manipulated materialism to further explore its mediating role. This study adopted a moderation-of-process design (Spencer et al., 2005), whereby a contextual condition to interrupt (vs. not interrupt) the causal process hypothesized and explained how the independent variable relates to the dependent variable (i.e., how self-esteem affects the corrupt intention; Jacoby and Sassenberg, 2011). A stronger ground for mediation exists if the manipulated materialism meaningfully moderates the basic effect of interest. That is, if inducing materialism can attenuate the relationship between self-esteem and corrupt intention, we can conclude that it is materialism, not another variable, that accounts for the effect of self-esteem on corrupt intention.

STUDY 1: CORRELATED RESEARCH

In Study 1, we aimed to identify direct associative evidence for the possible relationship among self-esteem, materialism, and corrupt intention. We predicted that self-esteem would be negatively correlated with corrupt intention and that materialism would mediate this correlation.

Method

Ethics Statement

The study was reviewed and approved by the Committee of Protection of Subjects at Beijing Normal University. All participants provided written informed consent before the study and were debriefed at the end of the research according to the established committee guidelines. This procedure was also followed in Studies 2 and 3.

Participants

We recruited 462 participants (265 women, 197 men) from two universities in China. The participants were between the ages of 17 and 22 ($M = 18.73$ years, $SD = 1.11$). To ensure the diversity of the sample, we included participants from different majors such as biology, accounting, information technology, education, and the arts.

Materials

Self-esteem

The widely recognized Chinese version of the 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1965; Li et al., 2011) was used to measure participants' self-esteem. One example item was "I feel that I'm a person of worth, at least on an equal plane with others." The participants were instructed to indicate their agreement with each statement on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The self-esteem index was calculated as the average score of these 10 items, with higher scores representing higher self-esteem (*Cronbach's* $\alpha = 0.824$).

Materialism

The widely recognized Chinese version of the 18-item Material Values Scale (MVS; Richins and Dawson, 1992; Li and Guo, 2009) was used to measure participants' materialism. One example item was "I admire people who own expensive homes, cars, and clothes." The participants were instructed to indicate their agreement with each statement on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The materialism indicator was calculated as the average score of these 18 items, and higher scores represented a higher level of materialism (*Cronbach's* $\alpha = 0.802$).

Corrupt intention

A 14-item corrupt intention measure (Leong and Lin, 2009) was adapted to measure participants' corrupt intention. We changed a few wordings of the original scale to assess the personal intention to engage in corruptive behavior. For example, the original item "Business corruption is inevitable in some cultures" was adapted by deleting "in some cultures"; the original item "When dealing with a business partner from abroad, it is important to inform the relevant authorities if the overseas partner asks for a bribe (R)" was adapted by deleting "from abroad" and "overseas"; the original item "In some countries, it is alright to pay someone extra in order to get things done quickly even if the law forbids such practices" was adapted by using "In some situations" to replace "In some countries." The measure was translated into Chinese and back-translated for accuracy by a Chinese–English bilingual speaker. Participants were instructed to indicate their agreement with each statement on a 9-point Likert scale from 1 ("completely disagree") to 9 ("completely agree"). The average score of these 14 items was calculated as a corrupt intention index, with a higher rating representing a stronger corrupt intention (*Cronbach's* $\alpha = 0.773$).

Procedure

After providing informed consent, the participants completed an online survey, including several questionnaires, in a computer room. These questionnaires included the self-esteem scale, the materialism measure, the corrupt intention measure, and other unrelated measures to prevent the participants from guessing the purposes of the research. After the participants completed the questionnaires, they were asked to provide their demographic information, including their sex, age, major, and birthplace. Finally, the participants were debriefed.

Results and Discussion

Preliminary Analyses

The correlations between the three variables, the means, and the standard deviations are shown in **Table 1**. The results demonstrated that self-esteem was negatively associated with

TABLE 1 | Means, standard deviations, and correlation matrix among all variables.

Variables	Mean	*SD*	1	2
(1) Self-esteem	4.82	0.92		
(2) Materialism	3.61	0.81	−0.264**	
(3) Corrupt intention	4.06	1.16	−0.181**	0.434**

**$p < 0.01$.

corrupt intention ($r = -0.181$, $p < 0.001$), which supported Hypothesis 1.

Additionally, we observed that each participant's gender and age had nearly significant correlations with corrupt intention, which was consistent with previous research indicating that demographic variables, such as age and gender, are related to moral reasoning (Haidt et al., 1993) and corrupt intention (Kennedy and Kray, 2014; Tan et al., 2016b). Thus, we subsequently included the covariance paths for age and gender in our mediation analyses.

Mediation via Materialism

We then tested our prediction that materialism mediates the association between self-esteem and corrupt intention using bootstrapping procedures (Preacher and Hayes, 2004). The analyses showed that self-esteem significantly predicted corrupt intention ($\beta = -0.182$, $SE = 0.056$, $t = -4.127$, $p < 0.001$). The variations in materialism predicted by self-esteem (a; $\beta = -0.273$, $SE = 0.040$, $t = -6.048$, $p < 0.001$) and the variations in corrupt intention predicted by materialism (b; $\beta = 0.397$, $SE = 0.059$, $t = 9.582$, $p < 0.001$) were both significant. After controlling for the effect of materialism, the direct effect of self-esteem on corrupt intention became non-significant ($\beta = -0.080$, $SE = 0.054$, $t = -1.89$, $p = 0.060$). A bootstrapping procedure comprising 5,000 samples provided additional evidence that the 95% confidence interval for the direct effect of self-esteem was [−0.207, 0.004], including zero, whereas the indirect effect was [−0.198, −0.073], not including zero (see **Figure 1**). These results support Hypothesis 2 that materialism accounts completely for the association between self-esteem and corrupt intention.

STUDY 2: CAUSAL RESEARCH

Although Study 1 confirmed a negative association between self-esteem and corrupt intention, as well as the mediating role of materialism, it inadequately established a rigorous causal relationship. Potentially, high self-esteem decreases corrupt intention, but the opposite could also be plausible. To overcome this limitation, in Study 2, self-esteem was manipulated by random feedback on a personality test, thus placing participants into either a high or a control condition. We also extended the previous results by expanding on responses to a business scenario, which consisted of a behavior with real consequences. We expected that when self-esteem increased, individuals would exhibit lower materialism and a lower corrupt intention.

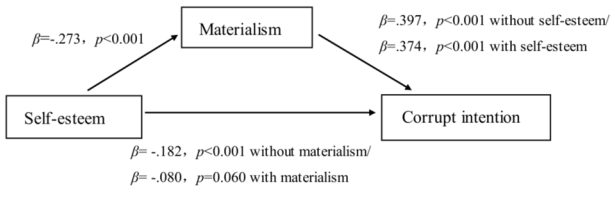

FIGURE 1 | Mediating model in Study 1 with controls including age and gender.

Method

Participants

One hundred participants (35 women, 64 men, 1 unreported) were recruited at a Chinese university through the campus network. The participants spanned the ages of 17 to 21 ($M = 19.09$ years, $SD = 0.84$). To ensure the diversity of the sample, participants from different majors were included. They were randomly assigned to one of two experimental conditions: the high self-esteem condition ($n = 44$) or the control condition ($n = 56$).

Materials

Manipulation of self-esteem

We adopted the research paradigm of Aronson and Mettee (1968) to manipulate self-esteem. The participants were asked to complete a shortened version of the California Personality Inventory (CPI) to evaluate their personalities. This version contained only 25 items from the six scales related to self-esteem. However, the primary experimental purpose of this test was merely to provide the opportunity and rationale for situationally manipulating the participants' self-esteem via pre-programmed feedback regarding the participants' personality test results. We emphasized that the computer would score the personality inventories through a complicating coding process and then provide feedback. The participants in the high self-esteem condition received the following feedback (Aronson and Mettee, 1968; Greenberg et al., 1992):

> "The profile indicates you have a stable personality and are not given to pronounced mood fluctuations of excitement or depression. Your stableness does not seem to reflect compulsive tendencies, but rather an ability to remain calm and level-headed in almost any circumstance. You are straightforward when making a decision, and never punctilious. You have strong heart. Any negative evaluations from others cannot threaten your sense of value. You are mentally mature for your age, and remain so going forward."

Participants in the control condition received the following feedback:

> "The profile indicates you have a fairly stable personality, but occasionally experience mood fluctuations of excitement or depression. Your stableness reflects your aspirations for freedom and independence from others, but it may be a bit unrealistic. Your profile suggests that you might be very careful but meticulous when making unimportant decisions. You have fairly strong heart, but you need to be more mature going forward."

To check the effectiveness of the manipulation, the participants were presented with a 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1965), as used in Study 1, on the computer (*Cronbach's* $\alpha = 0.645$).

Materialism

We administered the same materialism scale as used in Study 1 to measure participants' level of materialism (*Cronbach's* $\alpha = 0.821$).

Corrupt intention

A business corruption scenario (Mazar and Aggarwal, 2011) was adapted to determine the participants' corrupt intentions. Participants were assumed the role of a sales agent who had to compete against two other firms to win a contract from an international buyer and earn a commission. The sales agent was contemplating whether to offer an unofficial payment (bribe) to the potential international buyer to help win this contract. It was translated into Chinese and back-translated for accuracy by a Chinese–English bilingual speaker. After reading the scenario, the participants were asked to answer the following five questions (Huang et al., 2015): "As for me, the way of giving the money is not in my mind (R)," "If not taking other factors into consideration, I am willing to give the money," "I never consider giving the money (R)," "If I have the same situation to face in the future, I will still give the money," and "I think I will give the money to him." The responses to these items were scored on a 9-point Likert scale from 1 ("completely disagree") to 9 ("completely agree"). The average score of these five items was calculated as a corrupt intention index, with higher ratings representing higher corrupt intentions (*Cronbach's* $\alpha = 0.791$).

Procedure

All participants entered the laboratory and were informed that they were participating in a study concerned with the correlation between personality test scores and social behavior. The participants initially completed the CPI to evaluate their personalities and received random feedback, which randomly divided them into two conditions: the high self-esteem condition or the control condition. After receiving the random feedback (positive or neutral) regarding their personalities, the participants were instructed to complete the next questionnaire. At the end of the study, the participants were thanked with stationery gifts and debriefed.

Results and Discussion

Manipulation Check

As expected, participants in the high self-esteem condition had significantly higher self-esteem scores ($M = 4.76$, $SD = 0.59$) than those in the control condition did ($M = 4.48$, $SD = 0.68$), $t(98) = 2.18$, $p < 0.05$, *Cohen's* $D = 0.28$. This finding suggests that the manipulation was effective.

Preliminary Analyses

As expected, the manipulated self-esteem was significantly correlated with materialism ($r = -0.276$, $p < 0.01$) and corrupt intention ($r = -0.235$, $p < 0.05$). Furthermore, the mean materialism score of participants in the high self-esteem condition ($M = 4.47$, $SD = 1.02$) was significantly lower than that of the control condition ($M = 5.02$, $SD = 0.90$), $t(98) = -2.85$, $p < 0.01$, *Cohen's* $D = -0.57$. The mean corrupt intention score of the high self-esteem condition ($M = 3.80$, $SD = 1.44$) was significantly lower than that of the control condition ($M = 4.56$, $SD = 1.69$), $t(98) = -2.39$, $p < 0.05$, *Cohen's* $D = -0.49$. These results suggest that increased self-esteem had a negative effect on materialism and corrupt intention.

Mediation via Materialism

We then coded the high self-esteem and control conditions as +1 and 0, respectively, and further explored the mediating effect of materialism through bootstrapping procedures (Preacher and Hayes, 2004). The analyses showed that the self-esteem condition significantly predicted corrupt intention ($\beta = -0.235$, $SE = 0.319$, $t = -2.39$, $p = 0.019$). The variations in materialism predicted by the self-esteem condition (a; $\beta = -0.276$, $SE = 0.192$, $t = -2.85$, $p = 0.005$) and the variations in corrupt intention predicted by materialism (b; $\beta = 0.438$, $SE = 0.149$, $t = 4.823$, $p < 0.001$) were both significant. After controlling for the effect of materialism, the direct effect of the self-esteem condition on corrupt intention became non-significant ($\beta = -0.123$, $SE = 0.306$, $t = -1.31$, $p = 0.193$). A bootstrapping procedure with 5,000 samples provided additional evidence that the 95% *CI* for an indirect effect of self-esteem on corrupt intention through materialism was [−0.763, −0.125], not including zero. The direct effect was [−1.008, 0.207], including zero, which implies that materialism completely mediated the effect of self-esteem on corrupt intention.

The results were identical to the findings of the analysis with controlled variables including age and gender (see **Figure 2**). The analyses showed that the self-esteem condition significantly predicted corrupt intention ($\beta = -0.235$, $SE = 0.370$, $t = -2.091$, $p = 0.039$). The variations in materialism predicted by the self-esteem condition (a; $\beta = -0.296$, $SE = 0.216$, $t = -2.677$, $p = 0.009$) and the variations in corrupt intention predicted by materialism (b; $\beta = 0.423$, $SE = 0.161$, $t = 4.438$, $p < 0.001$) were both significant. After controlling for the effect of materialism, the direct effect of the self-esteem condition on corrupt intention became non-significant ($\beta = -0.118$, $SE = 0.356$, $t = -1.096$, $p = 0.276$). A bootstrapping procedure indicated that the 95% *CI* for an indirect association between the self-esteem condition and corrupt intention operating through materialism with the controlled variables

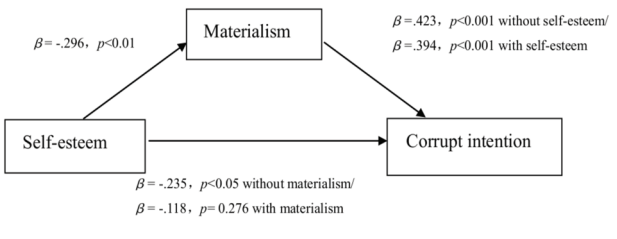

FIGURE 2 | Mediating model in Study 2 with controls including age and gender.

was [−0.818, −0.096], not including zero. The direct effect was [−1.099, 0.317], including zero, which implies materialism completely mediated the effect of self-esteem on corrupt intention.

The results showed that high self-esteem decreased the corrupt intention and that the buffered materialism mediated the relationship, thus further supporting our hypotheses. In other words, the results indicated that individuals whose self-esteem was enhanced would be less obsessed with material desires, thus were less likely to commit corrupt acts. Together, these results bolster our theoretical framework, thereby indicating that increasing self-esteem decreases the corrupt intention and that materialism helps explain this effect.

STUDY 3: PSYCHOLOGICAL PROCESS EXAMINING

Although Study 2 confirmed the causal link that high self-esteem decreased materialism and then buffered corrupt intention, the mediator accounting for the relationship between self-esteem and corrupt intention was essentially correlated. It is possible that evidence of mediation was obtained spuriously because of the relation between the measured mediator and the true psychological process. To confirm that the relationship between self-esteem and corrupt intention was influenced by materialism, in Study 3, we adopted the "moderation-of-process" design (Spencer et al., 2005) to further examine the psychological process. We primed materialism and predicted that the negative association between self-esteem and corrupt intention, as in Studies 1 and 2, would be reduced or eliminated when materialism was elicited. We expected that the differences in the degree of materialism could explain why individuals with varying self-esteem levels differ in their corrupt intention.

Method

Participants

A total of 127 participants (101 women, 25 men, 1 unreported) were recruited at a Chinese university. The participants spanned the ages of 18 to 27 (M = 20.75 years, SD = 2.29). They were randomly assigned to one of two experimental conditions: the materialism-induction condition (n = 63) or the control condition (n = 64).

Materials

Self-esteem

A 10-item Rosenberg Self-Esteem Scale was used, as in Study 1, to measure the participants' self-esteem levels (*Cronbach's* $\alpha = 0.817$).

Manipulation of materialism

To prime materialism, we relied on and adapted from a common experimental procedure, the scrambled-sentences task (Srull and Wyer, 1979; Bauer et al., 2012). The participants were presented with 30 word strings, each consisting of five words. For each string, the participants were instructed to select and order four of the words to form a valid Chinese sentence. For participants randomly assigned to the materialistic-cue condition, 20 of these word strings (67%) contained a word related to materialistic concepts (e.g., *buy, status, asset, and expensive*). For example, participants in this condition were asked to construct sentences out of such strings as "expensive, his, was, everybody, watch." In the control condition, highly similar word sets were created except that in each instance, materialistic concepts were replaced with mundane, non-materialistic ones (e.g., replacing the word *expensive* with the word *accurate*). Correspondingly, participants in this condition were presented with such neutral strings as "accurate, his, was, everybody, watch."

To check the effectiveness of the manipulation, the participants were asked to complete the identical materialism scale as in Study 1. We only changed prolonged words such as "always" in the original scale into present tense words such as "now" (*Cronbach's* $\alpha = 0.770$).

Corrupt intention

A business corruption scenario was used, as in Study 2, and five questions were asked to measure the participants' corrupt intention (*Cronbach's* $\alpha = 0.867$).

Procedure

After providing informed consent, the participants completed several experimental tasks on paper in a private cubicle. The first task was presented as a study of the "cognitive aspects of linguistic processing." Participants were asked to select and order four of the words to form a valid Chinese sentence. In reality, this was the priming task. Next, the participants completed ostensibly unrelated questionnaires. Finally, the participants were thanked, debriefed and paid RMB¥10 each.

Results and Discussion

Manipulation Check

As expected, the participants in the materialism-induction condition exhibited a higher materialistic tendency (M = 3.80, SD = 0.65) than the participants did in the control condition (M = 3.57, SD = 0.62), $t(125) = 2.06$, $p < 0.05$, *Cohen's* $D = 0.37$. This result suggests that the manipulation of materialism was effective. However, there was no difference in self-esteem between the materialism-induction (M = 4.97, SD = 0.82) and the control condition (M = 4.99, SD = 0.83), $t(125) = 0.10$, $p = 0.92$, suggesting that materialism did not influence self-esteem.

Test of Interaction

We predicted that the experimental condition would moderate the association between self-esteem and corrupt intention such that the negative relationship would be present in the control condition, but not in the materialism-induction condition. To test this prediction, we regressed corrupt intention with self-esteem, the experimental condition (materialism-induction vs. control), and the interaction of self-esteem and the experimental condition. To interpret the results, we centered self-esteem prior to the analysis because it is a continuous variable (Aiken et al., 1991).

TABLE 2 | Regression results predicting corrupt intention in Study 3.

	Model 1				Model 2			
Variable	***B***	***SE***	**β**	***t***	***B***	***SE***	**β**	***t***
Gender					−0.408	0.419	−0.084	−0.974
Age					0.090	0.075	0.106	1.204
Self-esteem	−0.611	0.195	−0.257	−3.14**	−0.680	0.215	−0.283	−3.16**
Condition (1 = materialism, −1 = control)	0.498	0.159	0.257	3.14**	0.509	0.160	0.261	3.19**
Self-esteem × Condition	0.487	0.195	1.265	2.497*	0.412	0.202	1.070	2.043*
	$R^2 = 0.176$ $F(3,123) = 8.780$**				$R^2 = 0.196$ $F(5,120) = 5.862$**			

*$p < 0.05$, **$p < 0.01$.

The results are presented in **Table 2**. After controlling for gender and age, self-esteem was negatively associated with corrupt intention. This association was qualified by a significant interaction between self-esteem and experimental condition ($p = 0.043$), displayed in **Figure 3**. To interpret the interaction, we tested the simple slopes using the procedures described by Aiken et al. (1991). In the control condition, the simple slope of self-esteem on corrupt intention was significant: *simple slope* $= -1.092$, $SE = 0.294$, $t(120) = -3.716$, $p < 0.001$, a finding consistent with Studies 1 and 2. By contrast, and consistent with our prediction, in the materialism-induction condition: *simple slope* $= -0.268$, $SE = 0.296$, $t(120) = -0.906$, $p = 0.367$. When primed to a materialistic mindset, the corrupt intention of participants with higher self-esteem was comparable to participants with lower self-esteem. This finding suggests that lower self-esteem individuals tended to favor corrupt behavior, at least partly because they experience a higher level of materialism than individuals with higher self-esteem do.

In Study 3, we adopted the moderation-of-process design and observed that the negative effect of self-esteem on corrupt intention observed in Studies 1 and 2 was reduced when materialism was experimentally induced. The results suggest that one reason that individuals with high self-esteem tend to have lower corrupt intentions is that they tend to have a lower passion for materialism. Further, corruption might serve as a coping pattern accompanied with high materialism to satisfy the needs of individuals lacking self-esteem. In sum, the results imply that increased self-esteem decreases corrupt intention through a lower level of materialism.

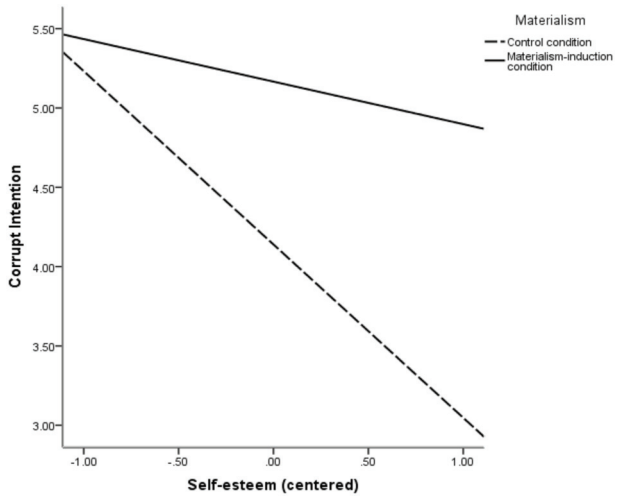

FIGURE 3 | Corrupt intention as a function of self-esteem and materialism in Study 3.

GENERAL DISCUSSION

The present three studies showed that high self-esteem decreased individuals' tendencies toward corruption and that materialism mediated this relationship based on a sample of Chinese university students. We confirmed our hypotheses in three studies using different measures of corrupt intention and different strategies to examine the process. The results of Study 1 suggest that self-esteem was negatively associated with the corrupt intention and that materialism mediated this relationship. In Study 2, we manipulated self-esteem, and the results indicated that increased self-esteem caused a low level of materialism and corrupt intention. In Study 3, we used a different strategy to examine the underlying psychological process, determining that the differences in the degree of materialism explain why individuals with varying self-esteem levels differ in their corrupt intention. These studies converged to show that increased self-esteem caused a low level of materialism, which in turn decreased corrupt intention.

The findings from the present research make a significant contribution to the literature of self-esteem. Previous research has shown that high self-esteem tends to be associated with, albeit might not cause, positive outcomes, such as better performance (Kuster et al., 2013), interpersonal success (Sommer and Baumeister, 2002), or health and well-being (Harter, 1999); whereas low self-esteem tends to be associated with problems such as cheat at a game (Aronson and Mettee, 1968), cheat on exams (Iyer and Eastman, 2006), and other dishonest behaviors (Graf, 1971). Our research expands the study of self-esteem to the field of corruption. To the best

of our knowledge, our results provide the first empirical evidence that high self-esteem decreases corrupt intention. Corruption impels individuals not only to compromise their ethical standards for their own benefit, but also to violate the law (Gupta et al., 2002; UN, 2003; Uslaner, 2008). Our findings imply that to maintain, enhance, and protect individuals' self-esteem is such a powerful motivation (Baumeister, 1998) that it drives individuals to commit corrupt acts with the heavy cost.

Furthermore, the present research also contributes to the effects of self-esteem on unethical behaviors like corruption, by identifying materialism as an underlying psychological process. The mediating role of materialism explains why individuals endorse less corruption after self-esteem is enhanced. On the one hand, the increasing self-esteem depresses materialism and corrupt intention. Material possessions are important for maintaining a self-concept (Belk, 1985) and are instruments for coping with doubts regarding self-worth or competence (Chang and Arkin, 2002; Jiang et al., 2015). Thus, when self-esteem is enhanced, the pursuit of material pleasures by means of corruption, which can temporarily produce prestige and wealth, is unnecessary, as inferred from the results of Study 2. On the other hand, the results indicate that materialism could buffer the negative relationship between self-esteem and corrupt intention. When primed to a materialistic mindset, the corrupt intention of individuals with higher self-esteem was comparable to individuals with lower self-esteem. Apart from coping with doubts concerning self-worth, materialism might also serve as a justification allowing individuals with higher self-esteem to commit corruption. This justification reduces the ethical dissonance, which represents the tension between moral-self and unethical behavior (Barkan et al., 2015). It could also be inferred from Study 3 that, if the pursuit of material decreases, individuals will be prone to commit less corrupt acts regardless of their self-esteem level, which would be very critical to control corruption.

The present research adds to the current debate on whether material possession amounts to self-worth. Previous research has found that an individual's self-esteem is negatively associated with materialism (Jiang et al., 2015), and that the relational change between the two variables over time has been demonstrated by further direct evidence (Yurchisin and Johnson, 2004; Chaplin and John, 2005; Isaksen and Roper, 2012). These results suggest that materialistic values could be used to cope with uncertainty about self-worth or competence (Chang and Arkin, 2002; Kasser, 2002; Jiang et al., 2015), thus might buffer threatened self-esteem. However, this function has only been observed as a temporary method to cope with the suffering of low self-esteem and might actually reduce an individual's well-being in the long term (Burroughs and Rindfleisch, 2002; Kasser et al., 2014). In fact, possession cannot amount to self-worth because increasing self-esteem increases subjective well-being and life satisfaction (Diener et al., 1995; Schimmack and Diener, 2003). In contrast, an over-emphasis of materialistic goals might augment negative emotions and depressive symptomatology (Kasser and Ryan, 1993; Kashdan and Breen, 2007), inhibit positive emotion and positive social relationships, hinder socialization, inflict losses on subjective well-being, and undermine life satisfaction (Diener and Biswas-Diener, 2002; Christopher et al., 2007; Jiang et al., 2015). In this regard, our Study 3 also showed that increasing materialism did not enhance self-esteem.

The results from the current research have practical implications for anti-corruption. For China, rapid economic growth has introduced multiple pressures and temptations, including status, riches, and fame. Lü (2000) shows that three common types of corruption exist concerning Chinese people's daily life: graft, rent-seeking, and prebendalism. The main thread linking these different types of corruption is that many individuals regard public office as a business (Van Klaveren, 1989; Tan et al., 2016a), at the expense of morality for status and possessions, which thus become measures of personal success and self-worth. Individuals with positive self-esteem may inhibit their sensitivity to external material possessions (Richins, 1999; Kasser and Kasser, 2001) and are less likely to engage in corrupt behaviors. For individuals lacking positive self-esteem, materialism may be controlled by removing materialistic messages regarding money, possessions, status, and image from the environment, by providing them with strategies to reduce the effect of those messages when they are encountered, or by exposing them to scenes or objectives that reflect nature (Weinstein et al., 2009; Kasser, 2016). Such suppression of materialism may decrease the incidence of corrupt behaviors. Additionally, by extensively exposing individuals to prevailing ethical norms (Gong and Wang, 2013; Tan et al., 2016a) and information on the potential long-term damage of materialism and corruption, they may determine that the costs likely outweigh the benefits and become less engrossed in corruption to achieve wealth.

Some limitations of the current research should be mentioned. First, it seems that receiving a positive feedback could potentially boost positive mood, and mood and emotion have been shown to be linked to ethics in some ways (Bazerman et al., 2011; Gino and Shea, 2012; Teper et al., 2015). However, we did not measure the participants' mood in Study 2. Future study should measure and thereby statistically control mood when manipulating self-esteem to further explore its effect on corruption. Second, also in Study 2, we explored the effect of self-esteem on corrupt intention by randomly assigning the participants to either high self-esteem condition or control condition. Unfortunately, we failed to successfully manipulate low self-esteem following the same paradigm of Aronson and Mettee (1968). Future study could use a different paradigm to manipulate low self-esteem for further exploring its effect. Third, only self-reported measures were used in current research to capture corrupt intention. However, past researchers have noted that the relationship between predictor variables and unethical intentions is often weaker than the relationship between predictor variables and actual unethical behavior (Kish-Gephart et al., 2010). Therefore, future study should further explore the effect of self-esteem on

corrupt behavior by using the method of bribery game (Abbink and Hennig-Schmidt, 2006; Huang et al., 2015).

AUTHOR CONTRIBUTIONS

YL contributed to all aspects of work for this article. LL contributed to conception, design, and revising the article critically. XT contributed to data collection, design, and interpretation. ZH and JD contributed to data analysis, interpretation, and revising the article critically. WZ contributed to interpretation and revising the article carefully.

FUNDING

The Natural Science Foundation of China (31571145). Beijing Social Science Foundation (13ZHB027). The Program of the Co-Construction with Beijing Municipal Commission of Education of China.

REFERENCES

Abbink, K., and Hennig-Schmidt, H. (2006). Neutral versus loaded instructions in a bribery experiment. *Exp. Econ.* 9, 103–121. doi: 10.1007/s10683-006-5385-z

Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.

Allport, G. W. (1955). *Becoming: Basic Considerations for a Psychology of Personality*. New Haven, CT: Yale University Press.

Aronson, E., and Mettee, D. R. (1968). Dishonest behavior as a function of differential levels of induced self-esteem. *J. Pers. Soc. Psychol.* 9, 121–127. doi: 10.1037/h0025853

Barkan, R., Ayal, S., and Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Curr. Opin. Psychol.* 6, 157–161. doi: 10.1016/j.appet.2014.04.003

Bauer, M. A., Wilkie, J. E., Kim, J. K., and Bodenhausen, G. V. (2012). Cuing consumerism situational materialism undermines personal and social well-being. *Psychol. Sci.* 23, 517–523. doi: 10.1177/0956797611429579

Baumeister, R. F. (1998). "The self," in *The Handbook of Social Psychology*, 4th Edn, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: McGraw-Hill), 680–740.

Baumeister, R. F., Heatherton, T. F., and Tice, D. M. (1993). When ego threats lead to self-regulation failure: negative consequences of high self-esteem. *J. Pers. Soc. Psychol.* 64, 141–156. doi: 10.1037/0022-3514.64.1.141

Baumeister, R. F., and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497–529. doi: 10.1037/0033-2909.117.3.497

Bazerman, M. H., Gino, F., Shu, L. L., and Tsay, C.-J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emot. Rev.* 3, 290–292. doi: 10.1177/1754073911402370

Belk, R. W. (1985). Materialism: trait aspects of living in the material world. *J. Consum. Res.* 12, 265–280. doi: 10.1086/208515

Berninghaus, S. K., Haller, S., Krüger, T., Neumann, T., Schosser, S., and Vogt, B. (2013). Risk attitude, beliefs, and information in a corruption game–An experimental analysis. *J. Econ. Psychol.* 34, 46–60. doi: 10.1016/j.joep.2012.11.004

Blaine, B., and Crocker, J. (1993). "Self-esteem and self-serving biases in reactions to positive and negative events: an integrative review," in *Self-Esteem: The Puzzle of Low Self-Regard*, ed. R. Baumeister (New York, NY: Plenum Press), 55–85.

Brown, J. D., and Marshall, M. A. (2001). Self-esteem and emotion: some thoughts about feelings. *Pers. Soc. Psychol. Bull.* 27, 575–584. doi: 10.1177/0146167201275006

Burroughs, J. E., and Rindfleisch, A. (2002). Materialism and well-being: a conflicting values perspective. *J. Consum. Res.* 29, 348–370. doi: 10.1086/344429

Chang, L., and Arkin, R. M. (2002). Materialism as an attempt to cope with uncertainty. *Psychol. Mark.* 19, 389–406. doi: 10.1002/mar.10016

Chaplin, L. N., and John, D. R. (2005). The development of self-brand connections in children and adolescents. *J. Consum. Res.* 32, 119–129. doi: 10.1086/426622

Charron, N. (2016). Do corruption measures have a perception problem? Assessing the relationship between experiences and perceptions of corruption among citizens and experts. *Eur. Polit. Sci. Rev.* 8, 147–171. doi: 10.1017/S1755773914000447

Chowdhury, R. M., and Fernando, M. (2013). The role of spiritual well-being and materialism in determining consumers' ethical beliefs: an empirical study with Australian consumers. *J. Bus. Ethics* 113, 61–79. doi: 10.1007/s10551-012-1282-x

Christopher, A. N., Drummond, K., Jones, J. R., Marek, P., and Therriault, K. M. (2006). Beliefs about one's own death, personal insecurity, and materialism. *Pers. Individ. Dif.* 40, 441–451. doi: 10.1016/j.paid.2005.09.017

Christopher, A. N., Lasane, T. P., Troisi, J. D., and Park, L. E. (2007). Materialism, defensive and assertive self-presentational tactics, and life satisfaction. *J. Soc. Clin. Psychol.* 26, 1145–1162. doi: 10.1521/jscp.2007.26.10.1145

Clark, M. S., Greenberg, A., Hill, E., Lemay, E. P., Clark-Polner, E., and Roosth, D. (2011). Heightened interpersonal security diminishes the monetary value of possessions. *J. Exp. Soc. Psychol.* 47, 359–364. doi: 10.1016/j.jesp.2010.08.001

Crocker, J. (2002). The costs of seeking self–esteem. *J. Soc. Issues* 58, 597–615. doi: 10.1111/1540-4560.00279

Crocker, J., and Park, L. E. (2004). The costly pursuit of self-esteem. *Psychol. Bull.* 130, 392–414. doi: 10.1037/0033-2909.130.3.392

Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1080/08870440902783628

Diener, E., and Biswas-Diener, R. (2002). Will money increase subjective well-being? *Soc. Indic. Res.* 57, 119–169. doi: 10.1023/A:1014411319119

Diener, E., Diener, M., and Diener, C. (1995). Factors predicting the subjective well-being of nations. *J. Pers. Soc. Psychol.* 69, 851–864. doi: 10.1037/0022-3514.69.5.851

Dittmar, H. (2008). *Consumer Culture, Identity and Well-Being: The Search for the "Good Life" and the "Body Perfect"*. New York, NY: Psychology Press.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., and Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychol. Sci.* 16, 328–335. doi: 10.1111/j.0956-7976.2005.01535.x

Gino, F., and Mogilner, C. (2014). Time, money, and morality. *Psychol. Sci.* 25, 414–421. doi: 10.1177/0956797613506438

Gino, F., and Shea, C. (2012). *Deception in Negotiations: The Role of Emotions*. New York, NY: Oxford University Press.

Gong, T., and Wang, S. (2013). Indicators and implications of zero tolerance of corruption: the case of Hong Kong. *Soc. Indic. Res.* 112, 569–586. doi: 10.1007/s11205-012-0071-3

Graf, R. G. (1971). Induced self-esteem as a determinant of behavior. *J. Soc. Psychol.* 85, 213–217. doi: 10.1080/00224545.1971.9918570

Greenberg, J., Solomon, S., Pyszczynski, T., Rosenblatt, A., Burling, J., Lyon, D., et al. (1992). Why do people need self-esteem? Converging evidence that self-esteem serves an anxiety-buffering function. *J. Pers. Soc. Psychol.* 63, 913–922. doi: 10.1037/0022-3514.63.6.913

Gupta, S., Davoodi, H., and Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Econ. Gov.* 3, 23–45. doi: 10.1007/s101010100039

Haidt, J., Koller, S. H., and Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *J. Pers. Soc. Psychol.* 65, 613–628. doi: 10.1037/0022-3514.65.4.613

Harter, S. (1999). *The Construction of the Self: A Developmental Perspective*. New York, NY: Guilford Press.

Huang, Z., Liu, L., Zheng, W., Tan, X., and Zhao, X. (2015). Walking the straight and narrow: the moderating effect of evaluation apprehension on the relationship between collectivism and corruption. *PLoS ONE* 10:e0123859. doi: 10.1371/journal.pone.0123859

Isaksen, K. J., and Roper, S. (2012). The commodification of self-esteem: branding and British teenagers. *Psychol. Mark.* 29, 117–135. doi: 10.1002/mar.20509

Iyer, R., and Eastman, J. K. (2006). Academic dishonesty: are business students different from other college students? *J. Educ. Bus.* 82, 101–110. doi: 10.1016/j.nedt.2009.03.001

Jacoby, J., and Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: testing process hypotheses by interaction. *Eur. J. Soc. Psychol.* 41, 180–190. doi: 10.1002/ejsp.762

Jiang, J., Zhang, Y., Ke, Y., Hawk, S. T., and Qiu, H. (2015). Can't buy me friendship? Peer rejection and adolescent materialism: implicit self-esteem as a mediator. *J. Exp. Soc. Psychol.* 58, 48–55. doi: 10.1016/j.jesp.2015.01.001

Johnston, M. (2012). Corruption control in the United States: law, values, and the political foundations of reform. *Int. Rev. Adm. Sci.* 78, 329–345. doi: 10.1177/0020852312438782

Jordan, J., Leliveld, M. C., and Tenbrunsel, A. E. (2015). The moral self-image scale: measuring and understanding the malleability of the moral self. *Front. Psychol.* 6:1878. doi: 10.3389/fpsyg.2015.01878

Kashdan, T. B., and Breen, W. E. (2007). Materialism and diminished well-being: experiential avoidance as a mediating mechanism. *J. Soc. Clin. Psychol.* 26, 521–539. doi: 10.1521/jscp.2007.26.5.521

Kasser, T. (2002). *The High Price of Materialism*. Cambridge, MA: MIT Press.

Kasser, T. (2016). Materialistic values and goals. *Annu. Rev. Psychol.* 67, 489–514. doi: 10.1146/annurev-psych-122414-033344

Kasser, T., and Kasser, V. G. (2001). The dreams of people high and low in materialism. *J. Econ. Psychol.* 22, 693–719. doi: 10.1016/S0167-4870(01)00055-1

Kasser, T., Rosenblum, K. L., Sameroff, A. J., Deci, E. L., Niemiec, C. P., Ryan, R. M., et al. (2014). Changes in materialism, changes in psychological well-being: evidence from three longitudinal studies and an intervention experiment. *Motiv. Emot.* 38, 1–22. doi: 10.1007/s11031-013-9371-4

Kasser, T., and Ryan, R. M. (1993). A dark side of the American dream: correlates of financial success as a central life aspiration. *J. Pers. Soc. Psychol.* 65, 410–422. doi: 10.1037/0022-3514.65.2.410

Kennedy, J. A., and Kray, L. J. (2014). Who is willing to sacrifice ethical values for money and social status? Gender differences in reactions to ethical compromises. *Soc. Psychol. Pers. Sci.* 5, 52–59. doi: 10.1177/1948550613482987

Kish-Gephart, J. J., Harrison, D. A., and Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethical decisions at work. *J. Appl. Psychol.* 95, 1–31. doi: 10.1037/a0017103

Kouchaki, M., Smith-Crowe, K., Brief, A. P., and Sousa, C. (2013). Seeing green: mere exposure to money triggers a business decision frame and unethical outcomes. *Organ. Behav. Hum. Decis. Process.* 121, 53–61. doi: 10.1016/j.obhdp.2012.12.002

Kuster, F., Orth, U., and Meier, L. L. (2013). High self-esteem prospectively predicts better work conditions and outcomes. *Soc. Psychol. Pers. Sci.* 4, 668–675. doi: 10.1177/1948550613479806

Ledeneva, A. V. (1998). *Russia's Economy of Favours: Blat, Networking and Informal Exchange*. Cambridge: Cambridge University Press.

Leong, C., and Lin, W. (2009). "Show me the money!Construct and predictive validation of the intercultural business corruptibility scale (IBCS)," in *Intercultural Relations in Asia*, ed. C.-H. L. J. Berry (Singapore: World Scientific), 151–176.

Li, H., Yang, J., Jia, L., and Zhang, Q. (2011). Attentional bias in individuals with different level of self-esteem. *Acta Psychol. Sin.* 43, 907–916. doi: 10.1089/cyber.2012.0223

Li, J., and Guo, Y. (2009). Revision of material valve scale in Chinese college students. *Stud. Psychol. Behav.* 7, 280–283.

Li, S., Triandis, H. C., and Yu, Y. (2006). Cultural orientation and corruption. *Ethics Behav.* 16, 199–215. doi: 10.1207/s15327019eb1603_2

Lü, X. (2000). *Cadres and Corruption: The Organizational Involution of the Chinese Communist Party*. Stanford, CA: Stanford University Press.

Magee, J. C., and Galinsky, A. D. (2008). Social hierarchy: the self-reinforcing nature of power and status. *Acad. Manag. Ann.* 2, 351–398. doi: 10.1080/19416520802211628

Maslow, A. H. (1968). *Toward a Psychology of Being*. NewYork, NY: Van Reinhold.

Mazar, N., and Aggarwal, P. (2011). Greasing the palm can collectivism promote bribery? *Psychol. Sci.* 22, 843–848. doi: 10.1177/0956797611412389

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

McGregor, I., Zanna, M. P., Holmes, J. G., and Spencer, S. J. (2001). Compensatory conviction in the face of personal uncertainty: going to extremes and being oneself. *J. Pers. Soc. Psychol.* 80, 472–488. doi: 10.1037/0022-3514.80.3.472

Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S. P., and Joseph, J. (2010). Emotional intelligence, individual ethicality, and perceptions that unethical behavior facilitates success. *Revista Psicol. Trabajo Organ.* 26, 35–45. doi: 10.5093/tr2010v26n1a3

Mogilner, C., and Aaker, J. (2009). "The time vs. money effect": shifting product attitudes and decisions through personal connection. *J. Consum. Res.* 36, 277–291. doi: 10.1086/597161

Mruk, C. (1995). *Self-Esteem: Research, Theory, and Practice*. New York, NY: Springer Publishing Company.

Noguti, V., and Bokeyar, A. L. (2014). Who am I? The relationship between self-concept uncertainty and materialism. *Int. J. Psychol.* 49, 323–333. doi: 10.1002/ijop.12031

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *J. Public Econ.* 93, 950–964. doi: 10.1016/j.jpubeco.2009.03.001

Organization for Economic Co-operation and Development [OECD] (2014). *Background Brief: The Rationale for Fighting Corruption*. Available at: http://www.oecd.org/cleangovbiz/49693613.pdf

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* 36, 717–731. doi: 10.3758/BF03206553

Pyszczynski, T., Greenberg, J., Solomon, S., Arndt, J., and Schimel, J. (2004). Why do people need self-esteem? A theoretical and empirical review. *Psychol. Bull.* 130, 435–468. doi: 10.1037/0033-2909.130.3.435

Richins, M. L. (1999). "Possessions, materialism, and other-directedness in the expression of self," in *Consumer Value: A Framework for Analysis and Research*, ed. M. B. Holbrook (London: Routledge), 85–104.

Richins, M. L., and Dawson, S. (1992). A consumer values orientation for materialism and its measurement: scale development and validation. *J. Consum. Res.* 19, 303–316. doi: 10.1086/209304

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Schimmack, U., and Diener, E. (2003). Predictive validity of explicit and implicit self-esteem for subjective well-being. *J. Res. Pers.* 37, 100–106. doi: 10.1016/S0092-6566(02)00532-9

Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. *Adv. Exp. Soc. Psychol.* 25, 1–65. doi: 10.1016/S0065-2601(08)60281-6

Sheldon, K. M., and McGregor, H. A. (2000). Extrinsic value orientation and "The tragedy of the commons." *J. Pers.* 68, 383–411. doi: 10.1111/1467-6494.00101

Shrum, L., Wong, N., Arif, F., Chugani, S. K., Gunz, A., Lowrey, T. M., et al. (2013). Reconceptualizing materialism as identity goal pursuits: functions, processes, and consequences. *J. Bus. Res.* 66, 1179–1185. doi: 10.1016/j.jbusres.2012.08.010

Sivanathan, N., and Pettit, N. C. (2010). Protecting the self through consumption: status goods as affirmational commodities. *J. Exp. Soc. Psychol.* 46, 564–570. doi: 10.1016/j.jesp.2010.01.006

Solomon, S., Greenberg, J., and Pyszczynski, T. (1991). "A terror management theory of social behavior: the psychological functions of self-esteem and cultural worldviews," in *Advances in Experimental Social Psychology*, ed. M. Zanna (San Diego, CA: Academic Press), 93–159.

Sommer, K. L., and Baumeister, R. F. (2002). Self-evaluation, persistence, and performance following implicit rejection: the role of trait self-esteem. *Pers. Soc. Psychol. Bull.* 28, 926–938. doi: 10.1177/014672020280007006

Spencer, S. J., Zanna, M. P., and Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *J. Pers. Soc. Psychol.* 89, 845–851. doi: 10.1037/0022-3514.89.6.845

Srull, T. K., and Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: some determinants and implications. *J. Pers. Soc. Psychol.* 37, 1660–1672. doi: 10.1037/0022-3514.37.10.1660

Steele, C. M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Adv. Exp. Soc. Psychol.* 21, 261–302. doi: 10.1016/S0065-2601(08)60229-4

Stockemer, D., LaMontagne, B., and Scruggs, L. (2013). Bribes and ballots: the impact of corruption on voter turnout in democracies. *Int. Polit. Sci. Rev.* 34, 74–90. doi: 10.1177/0192512111419824

Tan, X., Liu, L., Huang, Z., Zhao, X., and Zheng, W. (2016a). The dampening effect of social dominance orientation on awareness of corruption: moral outrage as a mediator. *Soc. Indic. Res.* 125, 89–102. doi: 10.1007/s11205-014-0838-9

Tan, X., Liu, L., Zheng, W., and Huang, Z. (2016b). Effects of social dominance orientation and right-wing authoritarianism on corrupt intention: the role of moral outrage. *Int. J. Psychol.* 51, 213–219. doi: 10.1002/ijop.12148

Tang, T. L.-P., Chen, Y.-J., and Sutarso, T. (2008). Bad apples in bad (business) barrels: the love of money, Machiavellianism, risk tolerance, and unethical behavior. *Manag. Decis.* 46, 243–263. doi: 10.1108/00251740810854140

Tang, T. L.-P., and Chiu, R. K. (2003). Income, money ethic, pay satisfaction, commitment, and unethical behavior: is the love of money the root of evil for Hong Kong employees? *J. Bus. Ethics* 46, 13–30. doi: 10.1023/A:1024731611490

Tang, T. L.-P., and Liu, H. (2011). Love of money and unethical behavior intention: does an authentic supervisor's personal integrity and character (ASPIRE) make a difference? *J. Bus. Ethics* 107, 295–312. doi: 10.1007/s10551-011-1040-5

Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210. doi: 10.1037/0033-2909.103.2.193

Teper, R., Zhong, C. B., and Inzlicht, M. (2015). How emotions shape moral behavior: some answers (and questions) for the field of moral psychology. *Soc. Pers. Psychol. Compass* 9, 1–14. doi: 10.1111/spc3.12154

Tice, D. M. (1993). "The social motivations of people with low self-esteem," in *Self-Esteem: The Puzzle of Low Self-Regard*, ed. R. Baumeister (New York, NY: Plenum), 37–54.

Treisman, D. (2000). The causes of corruption: a cross-national study. *J. Public Econ.* 76, 399–457. doi: 10.1016/S0047-2727(99)00092-4

UN (2003). *Resolution 58/4 of 31 October: United Nations Convention against Corruption*. Available at: http://www.un-documents.net/a58r4.htm

Uslaner, E. M. (2008). *Corruption, Inequality, and the Rule of Law: The Bulging Pocket Makes the Easy Life*. New York, NY: Cambridge University Press.

Van Klaveren, J. (1989). "The concept of corruption," in *Political Corruption: A Handbook*, eds A. J. Heidenheimer, M. Johnston, and V. T. LeVine (New Brunswick, NJ: Transaction Publishers), 89–91.

Vohs, K. D., and Heatherton, T. F. (2001). Self-Esteem and threats to self: implications for self-construals and interpersonal perceptions. *J. Pers. Soc. Psychol.* 81, 1103–1118. doi: 10.1037/0022-3514.81.6.1103

Vohs, K. D., Mead, N. L., and Goode, M. R. (2006). The psychological consequences of money. *Science* 314, 1154–1156. doi: 10.1126/science.1132491

Vohs, K. D., Mead, N. L., and Goode, M. R. (2008). Merely activating the concept of money changes personal and interpersonal behavior. *Curr. Dir. Psychol. Sci.* 17, 208–212. doi: 10.1111/j.1467-8721.2008.00576.x

Wattanasuwan, K. (2005). The self and symbolic consumption. *J. Am. Acad. Bus.* 6, 179–184.

Weinstein, N., Przybylski, A. K., and Ryan, R. M. (2009). Can nature make us more caring? Effects of immersion in nature on intrinsic aspirations and generosity. *Pers. Soc. Psychol. Bull.* 35, 1315–1329. doi: 10.1177/0146167209341649

Yang, Q., Wu, X., Zhou, X., Mead, N. L., Vohs, K. D., and Baumeister, R. F. (2013). Diverging effects of clean versus dirty money on attitudes, values, and interpersonal behavior. *J. Pers. Soc. Psychol.* 104, 473–489. doi: 10.1037/a0030596

You, J.-S. (2007). Corruption as injustice. *Paper Presented at the Annual Meeting of the American Political Science Association*, Chicago, IL.

Yurchisin, J., and Johnson, K. K. (2004). Compulsive buying behavior and its relationship to perceived social status associated with buying, materialism, sel-esteem, and apparel-product involvement. *Fam. Consum. Sci. Res. J.* 32, 291–314. doi: 10.1177/1077727X03261178

Zhou, X., Vohs, K. D., and Baumeister, R. F. (2009). The symbolic power of money reminders of money alter social distress and physical pain. *Psychol. Sci.* 20, 700–706. doi: 10.1111/j.1467-9280.2009.02353.x

Zywica, J., and Danowski, J. (2008). The faces of Facebookers: investigating social enhancement and social compensation hypotheses; predicting Facebook^{TM} and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks. *J. Comput. Mediat. Commun.* 14, 1–34. doi: 10.1111/j.1083-6101.2008.01429.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 September 2016
doi: 10.3389/fpsyg.2016.01317

Why Does the "Sinner" Act Prosocially? The Mediating Role of Guilt and the Moderating Role of Moral Identity in Motivating Moral Cleansing

Wan Ding[1,2†], Ruibo Xie[1†], Binghai Sun[2], Weijian Li[2], Duo Wang[3] and Rui Zhen[1]*

[1] School of Psychology, Beijing Normal University, Beijing, China, [2] College of Teacher Education, Zhejiang Normal University, Jinhua, China, [3] School of Social Science, Policy and Evaluation, Claremont Graduate University, Claremont, CA, USA

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Ana M. Franco-Watkins,
Auburn University, USA
Rachel Barkan,
Ben-Gurion University of the Negev, Israel
Eyal Gamliel,
Ruppin Academic Center, Israel

***Correspondence:**
Binghai Sun
jky18@zjnu.cn

† These authors contributed equally to this study and share first authorship.

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *29 December 2015*
Accepted: *18 August 2016*
Published: *08 September 2016*

Citation:
Ding W, Xie R, Sun B, Li W, Wang D and Zhen R (2016) Why Does the "Sinner" Act Prosocially? The Mediating Role of Guilt and the Moderating Role of Moral Identity in Motivating Moral Cleansing. Front. Psychol. 7:1317. doi: 10.3389/fpsyg.2016.01317

Numerous studies have found that people tend to commit prosocial acts subsequent to previous immoral acts, as a response to the latter. This phenomenon is called moral cleansing or moral compensation. However, the specific mechanism how previous immoral acts motivate moral compensatory behaviors is still not fully understood. This study aimed to examine the roles of guilt and moral identity in the relation between previous immoral acts and subsequent prosocial behaviors to clarify the mechanism. Based on the extant research, the current study proposed a moderated mediation model to illustrate the process of moral cleansing. Specifically, a previous immoral act motivates guilt, which further leads to subsequent prosocial behaviors, while moral identity facilitates this process. The participants were primed by a recalling task (immoral act vs. a neutral event). The results support the hypothesized model and provide a framework that explains moral cleansing by integrating the roles of guilt and moral identity. These findings highlight the dynamic nature of people's morality with regard to how people adapt moral behaviors to protect their moral self-image.

Keywords: moral cleansing, moral compensation, ethical dissonance, moral identity, guilt, moral self-image

INTRODUCTION

Past research shows us that people behave dishonestly, but at the same time manage to perceive themselves as good and honest (Mazar et al., 2008; Jordan et al., 2011; Shalvi et al., 2015). According to the common theoretical model of self-maintenance, people are torn between their wish to be moral and the temptation to profit from dishonesty. This conflict is termed ethical dissonance (Barkan et al., 2012, 2015). Ethical dissonance elicits intense psychological tension and poses a threat to people's self-concept and well-being (Mulder and Aquino, 2013; Barkan et al., 2015). To reduce the tension and maintain the moral self, people use justification mechanisms. Some justifications can take place before people commit the ethical violation, such as ambiguity of rules, the prosocial nature of the act, and moral licensing (Shalvi et al., 2015). They enable people to excuse misbehaviors as less immoral and thus reduce anticipated ethical dissonance. More often, other justifications emerge after people's moral misconduct, in order to minimize the experienced dissonance, wipe out feelings of guilt, and cleanse the self.

Moral cleansing, also known as moral compensation, is a set of compensating moral or worthy actions that cancels out the ethical violation that preceded it, allowing a person to turn a new leaf (Zhong and Liljenquist, 2006; Sachdeva et al., 2009). This effect has been largely documented in previous research and shows the inconsistency in morality (Jordan et al., 2011; Conway and Peetz, 2012; Gino et al., 2015). For instance, it was showed that people who recollected their past immoral behaviors showed high enthusiasm for prosocial activities and less dishonesty than those who recalled their past moral behaviors (Jordan et al., 2011; Gino et al., 2015). Conway and Peetz (2012) also found that participants who were asked to write their own past immoral behaviors reported stronger willingness to offer help and donate more money to charity. Since moral cleansing has been demonstrated by various studies, it is important to explore why "sinners" act prosocially after their misdeeds. In this context, we need to consider two key factors.

GUILT AND MORAL CLEANSING

Guilt is one of the negative consequences of ethical dissonance (Tangney, 1990; Zhong and Liljenquist, 2006). It is a measurable aspect of the psychological tension of ethical dissonance (Tangney et al., 2007). Research has shown that guilt has a unique role in ethical dissonance (Gino et al., 2009; Sheikh and Janoff-Bulman, 2010; Xu et al., 2014). As the dissonance is more acute, the feelings of guilt increase, which may put more pressure on the individual to reduce the threat of ethical dissonance on the self. Cleansing can be an effective way to reduce that tension (Barkan et al., 2015). Thus, as people feel more guilt, their tendency for cleansing will increase. Specifically, we hypothesize that an immoral act motivates guilt, which further leads to subsequent prosocial behaviors.

MORAL IDENTITY AND MORAL CLEANSING

Interestingly, people can feel more or less guilt. Some people may feel very guilty because they took a newspaper without paying for it, while other people may do worse things like stealing thousands and feel less guilty. This indicates the individual differences in people's response to their past immoral act, consistently immoral act (e.g., stealing), or inconsistently compensatory act (e.g., donating money or helping others; Zhong and Liljenquist, 2006; Martens et al., 2010). Past research has shown that everyone wants to be moral and sees morality as an important part of their identity (Blasi, 1993). Fine-tuning this concept, Aquino and Reed (2002) referred to the centrality of moral identity to one's self-concept. They showed that, for some people, moral identity is an important and central part of their general identity, whereas for other people (who also see themselves as moral) this component is less central (Aquino and Reed, 2002; Aquino et al., 2009). We hypothesize that, as moral identity is more central to one's self-concept, he or she will be more susceptible to ethical dissonance, experience more psychological tension, and act more prosocially to minimize this distress.

Moral identity is defined as "a self-conception organized around a set of moral traits" and it reflects the self-importance of morality (Aquino and Reed, 2002). The trait-based definition stems from Blasi's (1984) thought that some moral traits (e.g., caring or helpful) may stand more centrally in one's self-concept than others (e.g., honest or generous). Aroused by the drive to maintain consistency between self-conception and action, moral identity contributes to motivate moral actions, as a self-regulation (Blasi, 2004; Aquino et al., 2009; Brooks et al., 2013). Specifically, people spontaneously compare their current actions with their moral self-conception, and once the comparative deficit or the threat to their moral self is detected, the psychological distress is generated (Barkan et al., 2015; West and Zhong, 2015). This suggests that the centrality of moral identity facilitates the evoking of guilt and compensatory behavior in ethical dissonance (Zhong et al., 2010). Further, Mulder and Aquino (2013) conducted an empirical research on the relationship between moral identity and moral cleansing. The results showed a facilitating effect of moral identity on moral compensatory behavior.

THE CURRENT STUDY

Based on the theoretical models and existing findings related to ethical dissonance, we presumed the process of moral cleansing by integrating the roles of guilt and moral identity simultaneously. Specifically, when committing an immoral or indecent act, people will experience stronger discrepancy between their behaviors and existing moral identity, which can elicit guilt (Jordan et al., 2011; Mulder and Aquino, 2013). The higher a person's moral identity is, the stronger guilt he/her feels. The desire to reduce the guilt can motivate them to engage in moral actions to protect their moral self-image, or in other words, to wash away their sins (West and Zhong, 2015).

Correspondingly, we proposed a moderated mediation model to illustrate the roles of guilt and moral identity in moral cleansing. To be specific, guilt mediates the relationship between previous immorality and moral compensatory acts, whereas moral identity plays a moderating role in this process. It is necessary to examine moral cleansing from a cross-cultural perspective before illustrating its process, and it has never been demonstrated among Chinese who grew up in an oriental cultural background. Thus, in this study we examined this assumption in a sample of Chinese young adults. Above all, this study aimed (1) to examine the moral cleansing effect in a Chinese sample, and (2) to clarify the roles of guilt and moral identity and their interplay in moral cleansing. The corresponding hypotheses were as follows:

Hypothesis 1: Previous immorality will motivate the tendency to offer help.
Hypothesis 2: Guilt will mediate the relationship between previous immorality and moral compensatory acts, and moral identity will play a positive moderating role in this process.

MATERIALS AND METHODS

Participants

In total, 360 Chinese adults participated in this online study via Sojump. They were provided a chance to win a raffle prize of ¥100 (about $15). On an average, these participants were aged 23.74 years (SD = 5.98 years), ranging from 18 to 38 years. Further, 169 participants (47%) were male and 210 participants (58%) were undergraduate students.

Design

To examine the moral cleansing effect, the participants were randomly distributed to different recalling tasks: recalling their own previous immoral acts for the primed group (n = 180) and recalling their own neutral acts for the unprimed group (n = 180). To further clarify the association among previous immoral behavior, guilt, moral identity, and moral compensatory behavior, the last three variables were measured and the immorality of previous immoral behavior was evaluated on a 4-point scale (0 = neutral, 1 = a little immoral, 2 = moderately immoral, and 3 = very immoral).

Measures

Moral Identity

The internalization subscale of the moral identity measure (Aquino and Reed, 2002) was used to assess the centrality of moral identity. This subscale of the two-dimensional instrument has been shown to tap into the degree to which moral traits are central to the self-concept (Aquino and Reed, 2002) and has been used in several studies on moral functioning (Aquino et al., 2009; Mulder and Aquino, 2013). The measure presents participants with a list of nine adjectives that might describe a person (generous, helpful, hardworking, honest, kind caring, compassionate, fair, and friendly) and then asks them to "visualize the kind of person who has these characteristics and imagine how that person would think, feel, and act." After being asked to think about someone who possesses these traits, participants were presented with the five items. Sample items included, "Being someone who has these characteristics is an important part of who I am," and "It would make me feel good to be a person who has these characteristics." Each of the items was answered on a 7-point Likert-type scale (1 = strongly disagree and 7 = strongly agree). Then, the five items were averaged to determine the moral identity score for each participant ($\alpha = 0.85$).

Priming Manipulation Using the Recalling Task

The priming manipulation used a procedure designed by Zhong and Liljenquist (2006), which had been used in several studies (Barkan et al., 2012; Mulder and Aquino, 2013; Jordan et al., 2015). At the beginning of the priming, all participants read instructions stating that the researchers were interested in exploring people's memory of daily life events. Then, the participants in the primed group was asked to recall one of their past unethical events and to describe any details, feelings, or emotions they experienced, while participants in the control group were asked to write down certain occurrences that had happened since a week ago until the present (Mulder and Aquino, 2013). We coded the immorality of the recalled acts by the method adapted from Jordan et al. (2011). According to Kaptein's (2008) definition, for immoral behavior: "violating significant (social) moral norms that are acceptable to the larger community," the immorality of the recalled acts was evaluated on a 4-point scale (0 = neutral, 1 = a little immoral, 2 = moderately immoral, 3 = very immoral). This method helped in understanding the association between previous immorality and compensatory behavior. The intraclass correlation coefficient (ICC = 0.84) showed a high initial interrater reliability; three coders discussed discrepancies to arrive at a consensus.

Guilt

At the end of the recall task, all participants were presented with the guilt scale (GS; Ding, 2015) to measure their current feelings of guilt, which was adapted from Tangney et al. (1996) and Lewis's (1971) measurements. The guilt scale consists of 16 items, with five items in the dimensions of realizing one's own error ($\alpha = 0.86$), six items in the dimension of feeling ($\alpha = 0.91$), and five items in the dimensions of behavior tendency ($\alpha = 0.83$). Respondents answered each question on a 7-point Likert scale (1 = strongly disagree and 7 = strongly agree). The Cronbach's alpha reliabilities indicated that the GS achieved optimal levels as per psychometric requirements.

The Tendency of Volunteering Behavior

A method revised from Schnall et al.'s (2010) measure was utilized to measure the tendency of volunteering behavior. Specifically, after completing the guilt scale, the participants were informed that the study had ended. Then, a window popped up to show the "Ask for help" situation. The window stated, "There is another survey for which we need your help, without any pay. Any amount of help would be greatly appreciated. You are free to decide whether you will be willing to help us and to choose the time you wish to spend on the survey before the survey starts." The time ranged from 0 to 120 min, at intervals of 10 min. According to Korsgaard et al.'s (2010) study, the experimenter should state that the participants (a) would receive no incentive for participating and (b) were not obligated to participate. As Korsgaard et al. (2010) explain, it is a valid measure to assess the participants' volunteering behavior. The given answers were encoded to 0–12 (ranging from *volunteering no time* to *2 h,* in 10-min increments), according to Oswald's (1996, 2002) method. Thus, the tendency toward volunteering behavior was measured and encoded.

Procedure

The Institutional Review Board of Zhejiang Normal University in China approved the protocol of the present study, including the consent procedure. We also obtained consent from our participants. All materials and measures were completed online, anonymously. At the beginning of the on-line survey, the moral identity of all participants was measured, after which, some filler questionnaires (about 30 items) unrelated to morality were filled. Then, the participants in the two groups were primed or controlled with different recalling tasks, respectively. Next, the

guilt scale was used to measure the guilt of all the participants. Last, all subjects participated in a test for participants' prosocial intentions in a simulated "Ask for help" situation. In total, nine participants were excluded for failing to recall their previous immoral acts in the primed group, and 13 participants were excluded for their invalid questionnaires (six in the primed group and seven in the unprimed group), which involved choosing the same, completely random, or contradicting options for the items.

RESULTS

Preliminary Analyses

Before testing our predictions, we described the study variables using means and standard deviations of the measures, which have been shown in **Table 1**. Then, the Pearson correlation analysis was conducted to explore the basic relationships between previous immorality, guilt, moral identity, and helping time. These results are also presented in **Table 1**, indicating that the compensatory prosocial behavior was positively correlated with previous immorality, guilt, and moral identity.

Hypothesis Testing

The chi-square test and one-way ANOVA were conducted to examine Hypothesis 1. First, the chi-square test showed that 66.86% of the participants in the primed group offered help, whereas only 34.27% of participants in the unprimed group did so ($\chi^2 = 14.992$, $p < 0.001$, $\Phi = 0.21$). Then, the results of the ANOVA showed a significantly different tendency for engaging in volunteering behavior between the primed and control groups (primed group: 3.89 ± 1.81; control group: 2.80 ± 1.64; $t = 5.71$, $p < 0.001$, Cohen's $d = 0.63$). As predicted, relative to the control group, recalling previous immoral behavior in the primed group motivated moral compensatory intentions. Before examining the hypothesized moderated mediation model, we tested the effect of previous immorality on subsequent prosocial behavior using a regression analysis. The result ($B = 0.25$, $SE = 0.04$, $p < 0.001$) indicated that every 1-unit increase in the previous immorality predicted a 0.25-unit increase in moral compensatory behavior.

To examine the association among previous immorality, guilt, moral identity, and compensatory behavior, we tested the moderated mediation model (Hypothesis 2) according to Muller et al.'s (2005) multiple regression analysis process with centered variables (Aiken et al., 1991). The regression analysis was conducted using the enter method. Bootstrap confidence intervals (CI) were computed for the regressions coefficients and 95% CI not containing 0 indicates significant results (Erceg-Hurn and Mirosevich, 2008). The results have been presented in **Table 2** and **Figure 1**.

First, previous immorality had a significantly indirect effect on subsequent prosocial behavior through guilt. To be specific, the path from previous immorality to guilt and the path from guilt to subsequent prosocial behavior were significant. The indirect effect (from previous immorality to the subsequent prosocial behavior through guilt) equaled 0.06, which was the product of 0.46 (the path from previous immorality to guilt) and 0.13 (the path from guilt to prosocial behavior). In addition, the direct effect of previous immorality on prosocial behavior reduced from 0.25 to 0.19 after guilt was added to the model. This indicated the partial mediating effect of guilt in the relationship between previous immorality and prosocial behavior, and the mediating effect made up 24% (0.06/0.25) of the total effect. Moreover, for the analysis on helping time, the adjusted R^2 increased from 0.22 to 0.35 when guilt was added in the third step. That is, guilt had an additional R^2 value of 13%.

Secondly, the moderating effect of moral identity was shown on the direct and indirect path of moral compensation. Specifically, the interaction of previous immorality and moral identity (PIMI) had a significant effect on guilt and prosocial behavior. To present the moderating role of moral identity in moral compensation, we plotted the two interactions in **Figures 2** and **3**, at different levels of previous immorality (0–3) and moral identity centrality (1 *SD* above and below the mean for high and low levels). **Figure 2** illustrates the effects of previous immorality on subsequent helping time while **Figure 3** shows the effects of previous immorality on guilt for high and low levels of moral identity centrality. As shown in **Figure 2**, there is a stronger moral cleansing effect for people who have a high centrality of moral identity (high-MI) than for people who have a low centrality of moral identity (low-MI) at Level 3 [$M_{\text{high-MI}} = 1.49$, $M_{\text{low-MI}} = 0.87$, $t_{(39)} = 1.97$, $p < 0.05$, Cohen's $d = 0.62$]. As presented in **Figure 3**, stronger guilt was elicited for high-MI people than for low-MI people at almost all levels of immorality [Level 1: $M_{\text{high-MI}} = 0.06$, $M_{\text{low-MI}} = -0.83$, $t_{(30)} = 2.54$, $p < 0.01$, Cohen's $d = 0.89$; Level 2: $M_{\text{high-MI}} = 0.64$, $M_{\text{low-MI}} = 0.24$, $t_{(48)} = 1.43$, $p < 0.10$, Cohen's $d = 0.40$; Level 3: $M_{\text{high-MI}} = 0.94$, $M_{\text{low-MI}} = 0.43$, $t_{(39)} = 1.70$, $p < 0.05$, Cohen's $d = 0.51$].

DISCUSSION

This study had revealed a moral cleansing effect among Chinese young adults. More importantly, the results supported the assumptive model and provided a framework for explaining moral compensation by integrating the roles of guilt and moral identity. Specifically, previous immorality elicits the feeling of guilt, which further motivates moral compensatory behavior to alleviate this psychological distress. Moral identity facilitates the process of moral cleansing directly or through eliciting strong guilt.

Moral compensation exists both in eastern and western cultures, indicating that moral compensation is a cross-cultural phenomenon. Specifically, compared with the percentages of participants (66.7 and 33.3%) who chose a cleansing product (i.e., an antiseptic wipe, versus a non-cleansing product, i.e., a pencil) in Zhong and Liljenquist's (2006) study, the present study revealed a similar percentage in the primed (66.86%) and unprimed (34.27%) groups in terms of the tendency to offer help. However, a difference was found between the percentage of those who offered help in the present study and of those who did so (73.9 and 40.9%) in Zhong and Liljenquist's (2006) study. These comparisons showed a similar effect of moral compensation and a different level of helping behavior in participants from different

TABLE 1 | Descriptive statistics and correlation matrix.

	Content of recalling task											
	Neutral (n = 173)						Immoral (n = 165)					
	M	*SD*	1	2	3	4	*M*	*SD*	1	2	3	4
1. Immorality	0	0	–				1.97	0.82	–			
2. Guilt	0.60	0.14	–	–			5.20	1.27	0.55**	–		
3. Moral identity	4.97	0.86	–	0.48**	–		5.04	0.89	0.08	0.52**	–	
4. Helping time	3.89	1.81	–	0.33**	0.19*	-	2.80	1.64	0.27**	0.30**	0.23**	–

*$p < 0.05$, **$p < 0.01$.

Morality of the recalled act ranges from 0 (neutral) to 3 (very immoral). Guilt scores range from 1 (strongly disagree e) to 7 (strongly agree). Moral identity scores range from 1 (completely disagree) to 7 (completely agree). Helping time ranges from 0 (no help) to 12 (120 min).

TABLE 2 | The regression results for moderated mediation model.

Predictor variables	The first step Helping time			The second step Guilt			The third step Helping time		
	B	*SE*	95%CI	*B*	*SE*	95%CI	*B*	*SE*	95%CI
X: Previous immorality (PI)	0.25	0.04	[0.16, 0.34]	0.46	0.03	[0.39, 0.53]	0.19	0.03	[0.12, 0.26]
Mo: Moral identity (MI)	0.08	0.04	[−0.02, 0.18]	0.09	0.03	[−0.01, 0.19]	0.07	0.04	[−0.01, 0.13]
XMo: PIMI	0.11	0.04	[0.03, 0.19]	0.14	0.03	[0.08, 0.20]	0.12	0.04	[0.04, 0.22]
Me: Guilt (G)	–	–		–	–		0.13	0.03	[0.05, 0.21]
MeMo: GMI	–	–		–	–		−0.05	0.03	[−0.10, 0.01]
Adj R^2	0.22			0.48			0.35		
F	19.21			144.18			23.47		

Mo, moderator variable; Me, mediator variable.

cultural backgrounds. Extending previous findings (Zhong and Liljenquist, 2006; Conway and Peetz, 2012), the present study also found the quantitative association that a 1-unit increase in the previous immorality predicted a 0.25-unit increase in moral compensatory behavior. The findings indicated that the higher level of previous immorality that people recalled, the more prosocial behavior they would commit subsequently.

Previous immorality motivates moral compensatory behavior through guilt. This finding extended our existing understanding of the effect of guilt (Tangney et al., 2007; Gino et al., 2009; Sheikh and Janoff-Bulman, 2010) and supported the guilt-motivation perspective of moral cleansing (Zhong and Liljenquist, 2006; Xu et al., 2014). Specifically, the mediation model showed two sub-processes of moral cleansing. That is, previous immorality firstly leads to a sense of guilt, and secondly, this feeling of psychological distress motivates moral compensatory behavior. Further, the more immoral the recalled action is, the stronger is the feeling of guilt, and the higher is the prosocial behavior that is elicited. Besides, previous immorality also had a direct effect on moral compensatory behavior after controlling for the role of guilt. This suggested a possibility that previous immorality motivates moral cleansing directly as well as through other psychological tension.

The centrality of moral identity can facilitate the direct process from previous immorality to moral compensatory behavior. The findings supported Zhong et al.'s (2010) assumption and were consistent with the results of Mulder and Aquino's (2013) research. Extending Mulder and Aquino's (2013) study, the present study examined the moderating role of moral identity in the process of moral cleansing, with the role of guilt considered. The results showed that moral identity could act as a moderator in the direct process from previous immorality to moral compensatory behavior, as presented in **Figure 1**. That is, as compared with a person with low centrality of moral identity (low-MI), a high-MI person (moral identity is more central to him/her) is more inclined to compensate for their previous immorality and subsequently act more prosocially. Additionally, we presumed that the significantly direct process from previous immorality to moral compensatory behavior might be attributed to some other moral emotions (e.g., shame) or to psychological tension. Then, the centrality of moral identity may promote moral cleansing through evoking a feeling of shame or distress, which needs further exploration in future research.

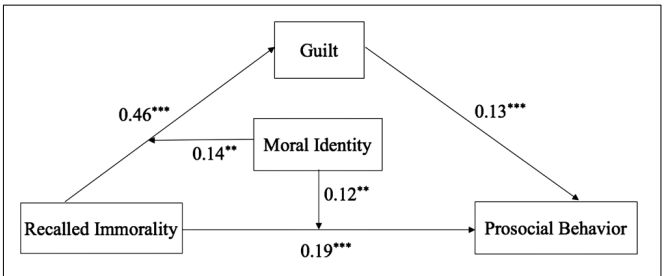

FIGURE 1 | The moderated mediation model. **$p < 0.01$, ***$p < 0.001$.

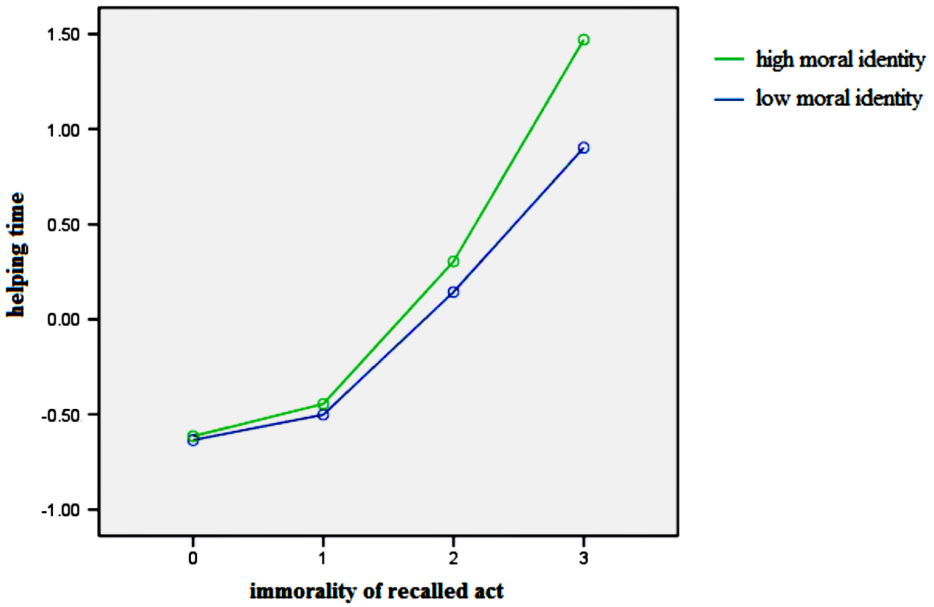

FIGURE 2 | The relationship between recalled immorality and helping time for high and low levels of moral identity.

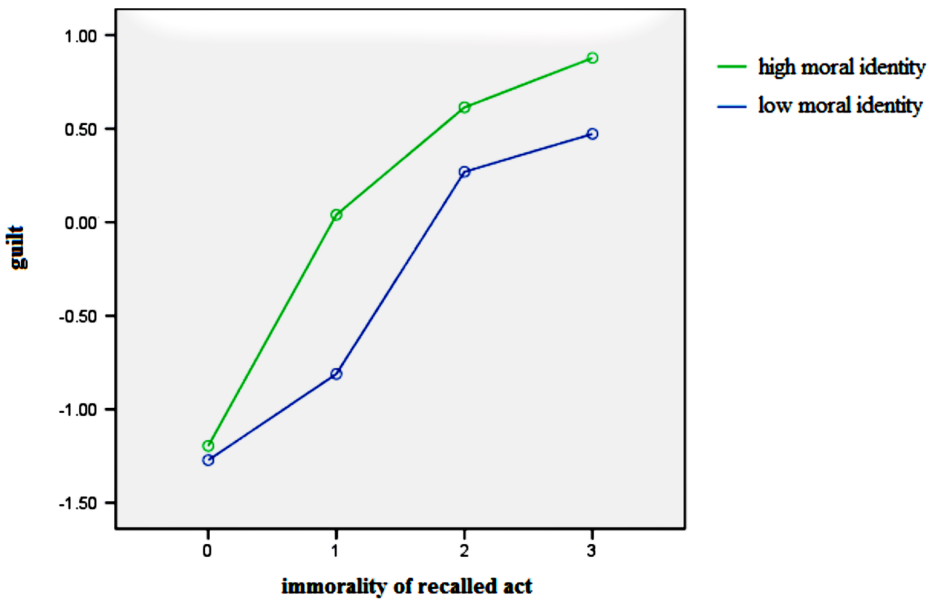

FIGURE 3 | The relationship between recalled immorality and guilt for high and low levels of moral identity.

More importantly, our finding on the interplay between guilt and moral identity helps to explain how guilt was elicited and influenced by moral identity in moral cleansing. It contributed to an adequate understanding of moral compensation. The results presented in **Table 2** and **Figure 3** indicated that the centrality of moral identity could facilitate the process from previous immorality to guilt. This finding supported the self-consistency theory (Barkan et al., 2015) and the self-comparison model of moral compensation (West and Zhong, 2015). When people recollect their own past immoral behavior, the inconsistency between one's self-conception and real conduct will lead to a sense of incompleteness and guilt. That is to say, they will feel guilt when they find themselves falling short of their existing moral identity. Thus, when moral identity is more important/central, the discrepancy between moral self-perception and the immorality of the recalled event is more pronounced, and people experience higher levels of guilt.

Interestingly, the demonstrated moderated mediation also contributes to explain moral consistency, in addition to moral inconsistency (e.g., moral compensation). The findings of the current study can help to explain why immoral behavior was followed by immoral behavior for some people (Zhong and Liljenquist, 2006; Martens et al., 2010). For low-MI people, their immoral act is consistent with their low moral identity. Therefore, it is possible that low-MI people mostly have no or just a low extent of discrepancy between immoral acts and moral identity, so they will experience less guilt or psychological distress. Low-MI people often continue their immoral behavior, but not moral compensatory behavior, after immoral acts.

Above all, combined with previous findings (Mulder and Aquino, 2013; Xu et al., 2014) and the self-consistency theory, the present study proposed and tested the moderated mediation model to show the mechanism underlying moral cleansing. The findings clarified and highlighted the vital importance of moral identity and guilt in moral self-regulation and the equilibrium of moral behavior. Specifically, previous immorality could not only motivate moral compensatory acts directly, but also through guilt. Besides, moral identity could facilitate the processes of evoking guilt and subsequent prosocial behavior by previous immorality. To sum up, the present study revealed that moral cleansing was observed among Chinese participants, and the findings showed us a framework to explain moral compensation with reference to the interplay between guilt and moral identity.

Several limitations should be addressed here. First, the present study only focused on the role of guilt, which may ignore the effects of other moral emotions in the process of moral compensation, such as shame. Therefore, other mediators should be distinguished in future studies. Second, the tendency of subsequent prosocial behavior was used to indicate participants' subsequent compensatory behavior. A gap between the tendency and actual behavior may affect the results. Hence, measures for the actual behavior should be considered in future studies. Third, the present study only focused on the internalization dimension of moral identity and did not put the role of the symbolization dimension (Jordan et al., 2011) into our consideration. Future research can integrate the two dimensions of moral identity to uncover the mechanism of moral cleansing.

Despite these limitations, to our knowledge, this is the first time to probe into the mechanism underlying moral compensation from a comprehensive perspective of combining guilt and moral identity. This study revealed a dynamic model on how people adapt moral behavior to protect their moral self-image. Furthermore, since the research was carried out on a Chinese population, it offers us a glimpse of the cross-cultural differences. Actually, it does not point at differences, but shows that despite the different cultural background, the same psychological processes of ethical dissonance and moral cleansing equally apply to Chinese participants. Finally, from the perspective of application, the present findings also have important implications for motivating the prosocial behaviors of "sinners." Practical efforts should concentrate on eliciting the discrepancy between sinners' desired state (moral identity) and their current state (recalling his own previous immorality) to induce their subsequent prosocial behaviors.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: WD, BS, RX, and WL. Performed the experiments: WD and RX. Analyzed the data: WD. Contributed to the writing of the manuscript: WD, RX, DW, and RZ.

FUNDING

This research was supported by grants from the National Social Science Foundation of China (CBA120107).

ACKNOWLEDGMENT

We are grateful to Dr. Xiuyun Lin for her help with this manuscript.

REFERENCES

Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks,CA: Sage.

Aquino, K., Freeman, D., Reed, A. II., Lim, V. K., and Felps, W. (2009). Testing a social-cognitive model of moral behavior: the interactive influence of situations and moral identity centrality. *J. Pers. Soc. Psychol.* 97, 123–141. doi: 10.1037/a0015406

Aquino, K., and Reed, A. II. (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83, 1423–1440. doi: 10.1037//0022-3514.83.6.1423

Barkan, R., Ayal, S., and Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Curr. Opin. Psychol.* 6, 157–161. doi: 10.1016/j.copsyc.2015.08.001

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *J. Exp. Psychol. Gen.* 141, 757–773. doi: 10.1037/a0027588

Blasi, A. (1984). "Moral identity: its role in moral functioning," in *Morality, Moral Behavior, and Moral Development*, eds W. Kurtines and J. Gewirtz (New York, NY: Wiley), 128–139.

Blasi, A. (1993). "The development of moral identity: Some implications for moral functioning," in *The Moral Self*, eds G. Noam and T. Wren (Cambridge, MA: MIT Press), 99–122.

Blasi, A. (2004). "Moral functioning: Moral understanding and personality," in *Moral Development, Self, and Identity*, eds D. Lapsley and D. Narvaez (Mahwah, NJ: Lawrence Erlbaum Associates), 335–347.

Brooks, J., Narvaez, D., and Bock, T. (2013). Moral motivation, moral judgment, and antisocial behavior. *J. Res. Character Educ.* 9, 149–165.

Conway, P., and Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Pers. Soc. Psychol. Bull.* 38, 907–919. doi: 10.1177/0146167212442394

Ding, W. (2015). *The Guilt of Chinese People and Its Role in Moral Compensation*. Master's thesis, Zhejiang Normal University, Jinhua.

Erceg-Hurn, D. M., and Mirosevich, V. M. (2008). Modern robust statistical methods an easy way to maximize the accuracy and power of your research. *Am. Psychol.* 63, 591–601. doi: 10.1037/0003-066X.63.7.591

Gino, F., Gu, J., and Zhong, C. B. (2009). Contagion or restitution? When bad apples can motivate ethical behavior. *J. Exp. Soc. Psychol.* 45, 1299–1302. doi: 10.1016/j.jesp.2009.07.014

Gino, F., Kouchaki, M., and Galinsky, A. D. (2015). The moral virtue of authenticity: how inauthenticity produces feelings of immorality and impurity. *Psychol. Sci.* 26, 1–14. doi: 10.1177/0956797615575277

Jordan, J., Leliveld, M. C., and Tenbrunsel, A. E. (2015). The moral self-image scale: measuring and understanding the malleability of the moral self. *Front. Psychol.* 6:1878. doi: 10.3389/fpsyg.2015.01878

Jordan, J., Mullen, E., and Murnighan, J. K. (2011). Striving for the moral self: the effects of recalling past moral actions on future moral behavior. *Pers. Soc. Psychol. Bull.* 37, 701–713. doi: 10.1177/0146167211400208

Kaptein, M. (2008). Developing a measure of unethical behavior in the workplace: a stakeholder perspective. *J. Manage.* 34, 978–1008. doi: 10.1177/0149206308318614

Korsgaard, M. A., Meglino, B. M., Lester, S. W., and Jeong, S. S. (2010). Paying you back or paying me forward: understanding rewarded and unrewarded organizational citizenship behavior. *J. Appl. Psychol.* 95, 277–290. doi: 10.1037/a0018137

Lewis, H. B. (1971). *Shame and Guilt in Neurosis*. New York, NY: International Universities Press.

Martens, A., Kosloff, S., and Jackson, L. E. (2010). Evidence that initial obedient killing fuels subsequent volitional killing beyond effects of practice. *Soc. Psychol. Pers. Sci.* 1, 268–273. doi: 10.1177/1948550609359813

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mulder, L. B., and Aquino, K. (2013). The role of moral identity in the aftermath of dishonesty. *Organ. Behav. Hum. Decis. Process.* 121, 219–230. doi: 10.1016/j.obhdp.2013.03.005

Muller, D., Judd, C. M., and Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *J. Pers. Soc. Psychol.* 89, 852–863. doi: 10.1037/0022-3514.89.6.852

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *J. Soc. Psychol.* 136, 613–623. doi: 10.1080/00224545.1996.9714045

Oswald, P. A. (2002). The interactive effects of affective demeanor, cognitive processes, and perspective-taking focus on helping behavior. *J. Soc. Psychol.* 142, 120–132. doi: 10.1080/00224540209603890

Sachdeva, S., Iliev, R., and Medin, D. L. (2009). Sinning saints and saintly sinners: the paradox of moral self-regulation. *Psychol. Sci.* 20, 523–528. doi: 10.1111/j.1467-9280.2009.02326.x

Schnall, S., Roper, J., and Fessler, D. M. (2010). Elevation leads to altruistic behavior. *Psychol. Sci.* 21, 315–320. doi: 10.1177/0956797609359882

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Sheikh, S., and Janoff-Bulman, R. (2010). The "shoulds" and "should nots" of moral emotions: a self-regulatory perspective on shame and guilt. *Pers. Soc. Psychol. Bull.* 36, 213–224. doi: 10.1177/0146167209356788

Tangney, J. P. (1990). Assessing individual differences in proneness to shame and guilt: development of the Self-Conscious Affect and Attribution Inventory. *J. Pers. Soc. Psychol.* 59, 102–111. doi: 10.1037/0022-3514.59.1.102

Tangney, J. P., Miller, R. S., Flicker, L., and Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *J. Pers. Soc. Psychol.* 70, 1256–1269. doi: 10.1037/0022-3514.70.6.1256

Tangney, J. P., Stuewig, J., and Mashek, D. J. (2007). "What's moral about the self-conscious emotions," in *The Self-conscious Emotions: Theory and Research*, eds J. L. Tracy, R. W. Robins, and J. P. Tangney (New York, NY: Guilford Press), 21–37.

West, C., and Zhong, C. B. (2015). Moral cleansing. *Curr. Opin. Psychol.* 6, 221–225. doi: 10.1016/j.copsyc.2015.09.022

Xu, H., Bègue, L., and Bushman, B. J. (2014). Washing the guilt away: effects of personal versus vicarious cleansing on guilty feelings and prosocial behavior. *Front. Hum. Neurosci.* 8:97. doi: 10.3389/fnhum.2014.00097

Zhong, C. B., Ku, G., Lount, R. B., and Murnighan, J. K. (2010). Compensatory ethics. *J. Bus. Ethics* 92, 323–339. doi: 10.1007/s10551-009-0161-6

Zhong, C. B., and Liljenquist, K. (2006). Washing away your sins: threatened morality and physical cleansing. *Science* 313, 1451–1452. doi: 10.1126/science.1130726

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 October 2015
doi: 10.3389/fpsyg.2015.01566

Binding lies

Avraham Merzel[1], Ilana Ritov[1,2], Yaakov Kareev[1,2] and Judith Avrahami[1,2]*

[1] School of Education, The Hebrew University of Jerusalem, Jerusalem, Israel, [2] The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem, Israel

Do we feel bound by our own misrepresentations? Does one act of cheating compel the cheater to make subsequent choices that maintain the false image even at a cost? To answer these questions we employed a two-task paradigm such that in the first task the participants could benefit from false reporting of private observations whereas in the second they could benefit from making a prediction in line with their actual, rather than their previously reported observations. Thus, for those participants who inflated their report during the first task, sticking with that report for the second task was likely to lead to a loss, whereas deviating from it would imply that they had lied. Data from three experiments (total $N = 116$) indicate that, having lied, participants were ready to suffer future loss rather than admit, even if implicitly, that they had lied.

Keywords: lies, binding, motivation, profit, commitment, self-presentation, impression management

INTRODUCTION

There are many reasons why people lie: to obtain material benefits, to impress, to save themselves from embarrassment or inconvenience, to avoid punishment, to protect a relationship, or even to benefit others (through white lies) (Hample, 1980; Camden et al., 1984; DePaulo et al., 1996; DePaulo and Kashy, 1998; Robinson et al., 1998; Vrij, 2000).

Although often beneficial, lies also bear some costs: lies violate the actual or perceived consistency, which is one of the foundations of interpersonal relationships (Cialdini et al., 1995). Lies degrade the quality of the information conveyed, thus diminishing the ability to arrive at an informed, high-quality decision (Lewicki, 1983). Lies impair interpersonal communication (Lewicki, 1983; Millar and Tesser, 1988; Grice, 1989). Lying entails internal psychological costs to the liar (Mazar and Ariely, 2006; Mazar et al., 2008). Finally, getting caught lying arouses negative emotions that affect both sides (Lewicki, 1983; Sagarin et al., 1998), and may result in actions (like punishment) against the liar (e.g., Lewicki, 1983; Mazar et al., 2008). It is thus obvious that people would be more likely to lie when they are not afraid of being exposed (Schlenker, 1975; Baumeister and Jones, 1978; Silverman et al., 1979; Baumeister, 1982; Leary and Kowalski, 1990; Mazar et al., 2008).

It is often the case that the behavior that follows lying may determine the likelihood of the lie being detected. It is therefore plausible that people would choose to act in a way that minimizes the chance of being exposed. But to what extent? Would people be ready to forgo a benefit in order that a future action does not reveal that they had previously lied?

In the present study, we examined whether, and to what extent, a person who lied is committed to the lie. Specifically, we wished to see if future actions made by that person would be affected by the commitment even at the cost of forgoing some profit.

We are not the first to study behavior that follows dishonest acts. For example, Both Mazar et al. (2008) and Chance et al. (2011) gave participants tasks that assessed their ability, while having the opportunity to dishonestly inflate their performance. Following these tests participants were

OPEN ACCESS

Edited by:
Guy Hochman,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Nina Mazar,
University of Toronto, Canada
Jason Dana,
University of Pennsylvania, USA

****Correspondence:***
Avraham Merzel
avraham.merzel@mail.huji.ac.il

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *22 June 2015*
Accepted: *28 September 2015*
Published: *14 October 2015*

Citation:
Merzel A, Ritov I, Kareev Y and Avrahami J (2015) Binding lies. Front. Psychol. 6:1566. doi: 10.3389/fpsyg.2015.01566

asked to predict their future performance in a similar task, this time without the opportunity to lie (for predictions of their past performance see Mazar and Hawkins, 2015). Their payment was determined both by their performance and by the accuracy of their prediction. Participants over-estimated their future performance. The authors interpret this over-estimation as reflecting self-deception. We wish to consider an alternative interpretation: impression management.

The study of lying behavior suffers from an inherent difficulty: on the one hand, to identify an action as a lie, one needs to observe actual behavior and compare it to the true state of affairs which is known, both by the potential liar and by the researcher. On the other hand, if participants are aware of being watched (i.e., that their lies could be exposed), the probability of them lying decreases (Mazar and Ariely, 2006).

One of the ways to deal with this dilemma is to set up a situation in which it is clear that no single lie can be caught. At the same time, liars can be identified with high probability by the degree to which the aggregate of their single, unverifiable, reports deviates from some reference value. Such a reference point can be based on the behavior of other people who could not lie when performing a similar task (Mazar and Ariely, 2006; Mazar et al., 2008; Chance et al., 2011; Schurr et al., 2012) or on statistical probability (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013; Hilbig and Hessler, 2013). In the present set of studies, we employed a paradigm of the latter type: the participants repeatedly performed a task in which they could falsely report a favorable outcome, with no fear of being caught. However, a comparison between the proportion of trials in which a favorable outcome was reported and the expected proportion of such outcomes could indicate whether the participants were likely to have lied. A subsequent task then tested how committed the participants were to their lies (if they lied). We describe our paradigm in the next section.

THE EXPERIMENTAL PARADIGM

The paradigm used here consisted of two tasks: a reporting task and a prediction task. In the reporting task the participants drew, privately, marbles from an opaque urn. The urn contained 100 marbles of two colors, and on each of 40 rounds the participants drew a marble and reported its color. One of the two colors was designated as the valuable color and reporting this color was rewarded. Thus in this task the participants had a financial incentive to inflate the number of marbles of the rewarding color, and did not run the risk of being caught lying. In the second task, which had not been announced in advance, the participants were asked to predict the number of marbles of the rewarding color that would show up in a sample of 40 marbles. Reward for performance in this task was based on the accuracy of the prediction. Obviously, the number that would most likely be accurate would be the number actually observed during the reporting task, rather than the number reported – if the latter had been inflated to obtain a larger reward. However, predicting a number that largely deviated from the value previously reported would constitute an admission, even if only tacit and indirect, that one had previously lied. Importantly, in our studies such a deviation would have no financial cost (the opposite: it would be financially rewarding). The only immaterial consequence would be that one would indirectly expose their lie. The question we explored was whether, and to what extent, participants whose reporting indicated they had most likely lied would be willing to suffer an expected loss, by predicting a value that was congruent with their previous report.

In what follows, we report three experiments that used this paradigm. Experiment 1 involved two phases, each consisting of a reporting and a predicting task, but with the second reporting task not incentivized. Experiment 2 used a preliminary, unincentivized, reporting task to rule out an alternative explanation based on anchoring. In Experiment 3 we introduced a procedure that guaranteed that reporting was anonymous to see if, when one's prediction could not be associated with one's previous reporting, participants would still be committed to their lies.

The experimental method employed in all three experiments reported was approved by the Human Subjects Committee of the School of Education, The Hebrew University of Jerusalem. All the participants signed an informed consent form before taking part in the experiments.

EXPERIMENT 1

The purpose of this experiment was to demonstrate that people are willing to forgo a possible profit in order to keep the false representation they displayed. It was further designed to address an alternative explanation based on the misperception of small probabilities.

Method

The experiment consisted of two, within-subject phases, each calling for the performance of the two tasks of reporting and predicting described above. The tasks of the second phase were identical to those of the first phase save for the fact that there was no incentive to lie during the reporting task. With, presumably, no lying in the latter reporting task the value in the prediction task was expected to correspond to the true proportion of marbles in the urn.

In the first phase the urn contained 35 green marbles and 65 yellow marbles and in the second phase it contained 35 blue marbles and 65 white marbles. The participants were informed that there were 100 marbles in the urns but they were not informed either of the number of marbles of each color or of the colors' ratio. In the reporting task, a laptop computer was located next to the urn. Both urn and computer were hidden from the experimenter by a curtain and were visible only to the participant. There were two keys on the computer screen, corresponding to the colors of the marbles in the urn. The participant was to draw a marble from the urn and report its color by clicking the corresponding key on the computer screen. The marble then had to be put back in the urn and the urn shuffled. The participants were instructed to repeat this procedure 40 times. Importantly, at that point the computer stopped and displayed

to the participant the number of the less common color (green in the incentivized reporting phase and blue in the unincentivized reporting phase), out of 40, that he or she had reported. At this stage the experimenter pulled the curtain aside, so the urn was visible to both the participant and the experimenter, and the predicting task began. The participant had to predict the number of marbles of the less common color in a sample of 40 marbles, drawn from the urn. After the prediction had been made and noted, the participant drew the marbles one at a time, and placed each marble in one of two separate containers, sorting the marbles by their color. Once 40 marbles had been drawn the experimenter and the participant counted the number of marbles of the relevant color together and compared it to the participant's prediction.

In the incentivized reporting task, every time the participant pressed the "green" key, the computer added 0.5 NIS (New Israeli Shekels, 1 NIS being worth 0.26 $ at the time of the experiment) to the participant's profits. It should be noted that the participants could report any color they wished without being exposed. The procedure of reporting in the second reporting phase was identical, but did not produce profit. This latter phase was introduced to make sure that the participants could correctly report the number of marbles of the infrequent color after 40 draws when there was no monetary incentive to inflate that number.

In the predicting task of both phases, payment was for *accuracy* in predicting the number of marbles of the infrequent color. The payment for a perfectly accurate prediction was an additional 5 NIS, and for a prediction that deviated by one from the number of marbles actually drawn it was 2 NIS. A prediction that deviated by more than one was not rewarded. Note that in the first phase this payment schedule could pose a dilemma for the participants who lied in the reporting task. On the one hand, their best prediction would have been the number of green marbles that they had *actually* drawn; on the other hand, if they were bound by their lies (i.e., lied and didn't want to get caught lying) they should predict the number of green marbles that they *reported* in the previous task. Clearly, in the latter case, those who inflated the number of green marbles could expect to forgo the reward for accuracy of prediction.

Participants

Thirty four students of the Hebrew University of Jerusalem were recruited for the experiment. The average age was 25.36 years ($SD = 3.42$ years) one student was excluded from the analysis, because he admitted that he didn't understand what was required of him. 14 of the 33 participants were females.

Procedure

Participants were tested individually in an empty classroom. The experimenter confirmed that the student could distinguish between the different colors of the marbles; those who had difficulty doing so were dismissed.

The instructions of the incentivized reporting task were read aloud and explained; written instructions were also available on paper in front of the participant. The same sequence of reading aloud and explaining, as well as providing written instructions, was true for all tasks.

After the instructions were explained, the participant went behind the curtain and started the task. Once the reporting task was over the prediction task started, with the participant predicting the number of green marbles that would be drawn out of 40 marbles. After that, the participant drew 40 marbles from the urn and sorted them by color into two boxes. The experimenter and the participant then counted the green marbles and the result was recorded. The second phase followed, starting with the reporting task and continuing with the prediction task for the blue/white urn. Finally, the total amount earned was calculated and paid, and the participant was dismissed.

Results

The mean number of the infrequent color, reported and predicted, for each phase and task, is presented in **Figure 1**. A two-way repeated measures ANOVA was conducted to test for the effects of Phase, Task, and their interaction. We found significant main effects of Phase, with the mean number higher in the first than in the second phase [$F(1,32) = 10.583$, $MSE = 37.867$, $p = 0.003$, $\eta_p^2 = 0.249$], and of Task, with the mean number higher in the reporting than in the predicting task [$F(1,32) = 6.145$, $MSE = 10.893$, $p = 0.019$, $\eta_p^2 = 0.161$]. The interaction was not significant [$F(1,32) = 0.950$, $MSE = 11.518$, $p = 0.337$, $\eta_p^2 = 0.029$).

While the results presented above capture the overall picture, it is easier to answer our research questions by reporting the results of pre-planned contrasts. First, the number of rewarding marbles reported in the first reporting task, in which misreporting was incentivized, was 19.55 – a value much higher than the value of 14, expected by chance [$t(32) = 4.67, p < 0.001$]. The prediction made following the incentivized reporting task, at 17.55, was also significantly larger than 14 [$t(32) = 4.36$, $p < 0.001$], indicating that the participants, although predicting a value somewhat lower than the one they had reported, were still committed to their lies. In fact, significant differences were observed not only between the reporting in the incentivized and unincentivized tasks [$t(32) = 3.070$, $p = 0.004$], but also, and most importantly, between the predictions made [$t(32) = 2.609$, $p = 0.014$]. The latter result indicates that the participants were ready to incur a loss in the predicting task to avoid having their prediction expose their lie.

As expected, the prediction following the unincentivized reporting task of the second phase (14.64) was not significantly higher than 14. However, the report in that phase (15.48) was, in fact, higher [$t(32) = 2.488$, $p = 0.018$]. We have no explanation for this result but it may be another indication of binding lies: participants may have suspected that the second reporting task could somehow expose their having lied before. Either way, the value of 15.48 is significantly different from either 19.55 [$p = 0.004$) or 17.55 ($p = 0.032$) in the incentivized phase.

To check if these findings may have resulted from a misperception of a difference in the statistical properties of the two procedures, reporting (calculating the expected proportion when drawing from an urn with replacement) and predicting

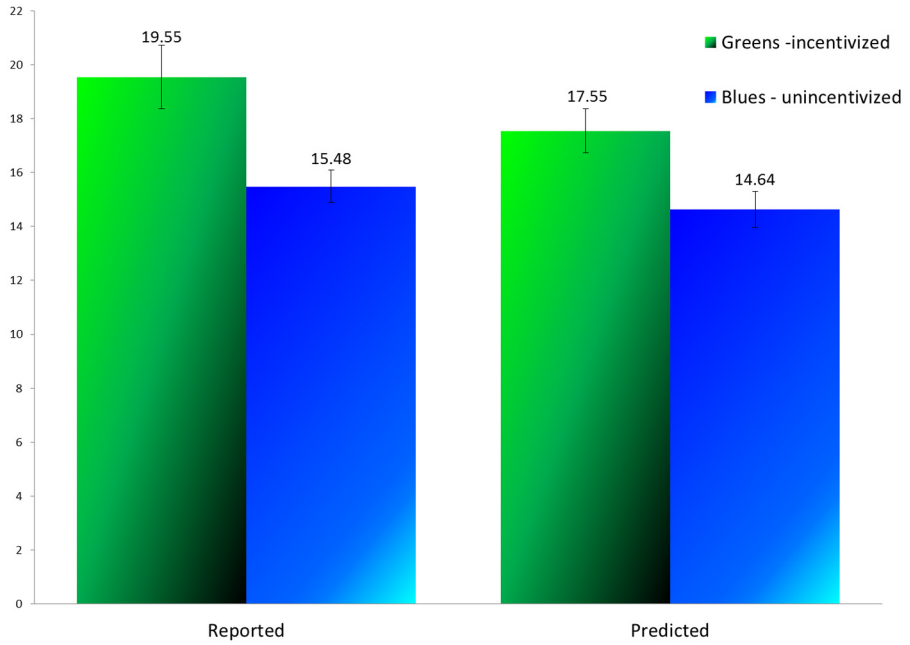

FIGURE 1 | Average of non-frequent marbles (reported or predicted) by phase (Experiment 1).

(calculating the expected proportion when drawing from an urn without replacement), we ran a control study with 39 subjects, none of whom had participated in the other study. In this study participants were presented with a written description of the procedure and results of the previously ran study, and asked what could have brought about these results. Only one of the 39 subjects gave an answer that could have been interpreted as referring to a difference between the statistical probabilities in the two procedures.

A comparison of earnings in the prediction tasks of the two phases (see **Figure 2**) revealed that earnings in the first phase (Mean Payment = 0.636 NIS, $SD = 1.517$) were indeed lower than that in the second phase [Mean Payment = 1.303 NIS, $SD = 1.811$; $t(32) = 1.785$, $p = 0.042$, one tailed]. In other words, had participants predicted what they must have *actually* observed while performing the incentivized reporting task they could have earned twice as much than they did. All in all, we conclude that the participants were willing to risk future gains rather than tacitly admit having lied before.

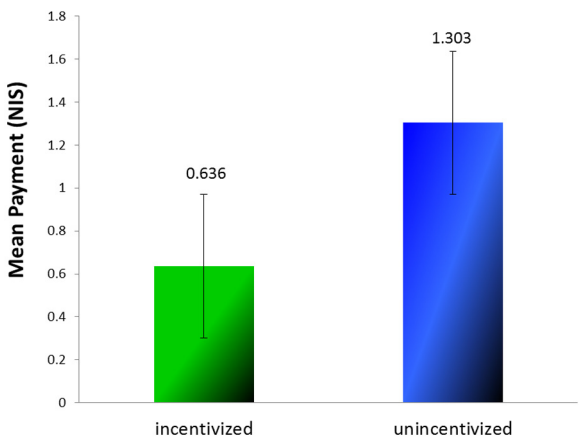

FIGURE 2 | Mean payment for prediction in the two phases (Experiment 1).

EXPERIMENT 2

It could be argued that the inflated number of green marbles reported (and then displayed) in the incentivized reporting task did not bind participants but served as an anchor for the next task. In other words, this claim means that the participants knew that while performing the reporting task they had retrieved fewer green marbles than they reported. It is therefore possible that, when they had to predict the number of green marbles that would be drawn; their correct estimate was drawn upward through anchoring, which resulted in an intermediate value.

In Experiment 2 we addressed the anchoring issue, while replicating the previous results.

Method

Experiment 2 was similar to Experiment 1, but included a preliminary, unincentivized reporting task, and no second phase.

In the preliminary reporting task the participants did exactly what they did later in the incentivized reporting task, but they did not get any money for reporting "green." For the preliminary task to make sense we asked the participants to use tongs to pull the marbles out of the urn – which was no simple matter – the unincentivized reporting task was described as "practice." At the end of the practice task the number of green marbles participants reported was presented to them on the computer screen. Following practice the participants performed the second, incentivized, reporting task followed by the prediction task – both identical to the tasks performed in Experiment 1 (except for the requirement to use tongs to draw marbles one by one). As before, in both reporting tasks it was possible for the participants to report that they had drawn a green marble even if it was of the other color. In this experiment the practice task served as a control in that it provided another possible anchor.

Participants

Thirty nine students of the Hebrew University of Jerusalem were recruited for the experiment. The average age was 25.64 years ($SD = 3.04$ years). Eighteen of the participants were females.

Procedure

The procedure was the same as the procedure of Experiment 1, except for the use of the tongs, the inclusion of a practice task, and the absence of a second phase.

Results

The average draws (reported and predicted) of the green marbles are presented for each task in **Figure 3**. We analyzed the results across all the participants with one-way repeated measures ANOVA. We found a significant difference between the three tasks [$F(2,76) = 9.398$, $MSE = 7.687$, $p < 0.001$, $\eta_p^2 = 0.198$]. The mean reported in the first, unincentivized task was closest to the expected value of 14, the mean reported in the second, incentivized, task was much higher, and the mean in the predicting task was somewhere in between. A finer contrast analysis that compared the mean number of green marbles in the predicting task separately to the mean reported in the practice task and to that reported in the incentivized task showed the prediction to differ significantly from both [$F(1,38) = 4.787$, $MSE = 12.340$, $p = 0.035$, $\eta_p^2 = 0.112$ and $F(1,38) = 6.050$, $MSE = 14.256$, $p = 0.019$, $\eta_p^2 = 0.137$, for the practice and the incentivized reporting tasks, respectively]. We examined the differences between every two tasks with paired-sample t-tests. All of the differences were significant: the difference between the practice task and the incentivized reporting task [$t(38) = 3.841$, $p < 0.001$], the difference between the incentivized reporting task and the prediction task [$t(38) = 2.46$, $p = 0.019$], and the difference between the prediction task and the practice task [$t(38) = 2.188$, $p = 0.035$].

These results replicate the results of Experiment 1: apparently at least some of the participants reported more green marbles than they had really drawn, that is to say, they lied. In the prediction task the participants predicted a lower number of green marbles than they had reported, but their prediction was not as low as what they had most likely seen in the two preceding reporting tasks.

Because the number of rewarding marbles observed by each subject is unknown, it is impossible to calculate how much profit participants had foregone by being bound by their lies. For approximation we calculated the mean payment the participants would have earned had they predicted the expected value (14). In that case they would have earned 1.513 NIS (see **Figure 4**); the difference between that and what they earned is significant [$t(38) = 1.842$, $p = 0.036$, one tailed, paired sample t-test].

The number reported during practice also differed from the expected value (14) [$t(38) = 2.555$, $p = 0.015$]. We have no explanation for this deviation, which could have resulted by chance. In any case, it does not bear on our main finding.

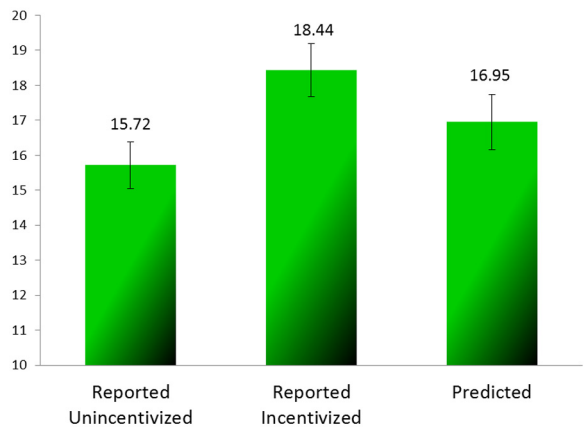

FIGURE 3 | Average of green marbles (reported or predicted) (Experiment 2).

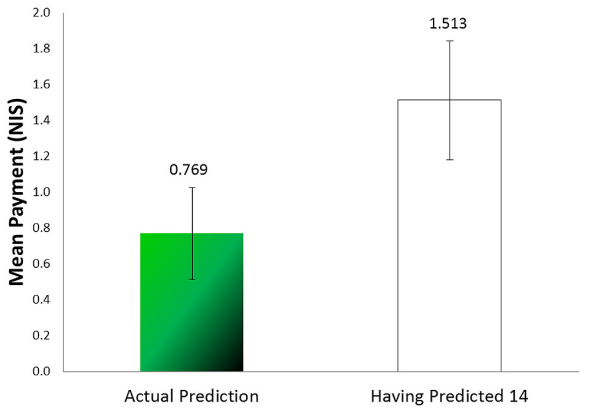

FIGURE 4 | Mean payment for prediction compared to a prediction of 14 (Experiment 2).

In Experiment 2 we not only replicated the findings of Experiment 1, but also ruled out an alternative explanation based on anchoring. The participants sampled the urn twice, and could anchor on either of the two values reported. The up deviation of the predicted value (16.95) from both the value expected by chance (14) and the value reported and displayed in the practice task is a clear indication that participants decided to stick to their inflated reports. At the same time, the difference between the reporting that was incentivized (18.44) and the subsequent prediction indicates not only that the participants were likely aware of the true proportion of green marbles in the urn, but also that they tried to "reduce" the damage caused by sticking to their inflated reports.

EXPERIMENT 3

Our main thesis is that behavior following a lie is affected by the lie in that the liar attempts to ensure that the act of lying is not exposed, even at a cost. Still, it could be claimed that what we regarded as the tell-tale indication of such an attempt – the large deviation of the prediction from the value expected by chance – resulted from other factors such as a failure to correctly estimate the proportion of the infrequent-color (green) marbles in the urn, self-deception (as would be predicted by the theory of self-concept maintenance Mazar et al., 2008; Chance et al., 2011), or still, by some anchoring. To test these alternative explanations we created in Experiment 3 a *non-binding* situation: not only could participants exaggerate their "report" of the number of rewarding marbles drawn with impunity, but also no one could tell how many rewarding marbles they reported. Thus, it would be impossible to find out if their prediction differed from their report, which would have implicated them as liars. We reasoned that under such conditions people would not be bound to their lies. On the other hand, if inflated predictions were the results of failures of estimation, self-deception, or anchoring they would remain higher than expected by chance, as was observed in the previous two experiments.

Method

This experiment was similar to Experiment 2 but with some changes: only in the practice stage did participants report into a computer. In the second stage, to allow participants to inflate their "reports," and still not be connected to that "report" (so that no one could tell if their prediction deviated from it), they had to count the number of draws and the number of green marbles for themselves. That way we could not tell, for any individual participant, how many green marbles she claimed to have drawn. Participants could use pen and paper but it was not obligatory. It was emphasized in the instructions that even if the participants were to use such aids, they would keep them and the experimenter would have no access to them.

At the end of the drawing the participant was presented with a large number of unmarked envelopes, each containing 40 0.50-shekel coins, and was asked to select one of the envelopes at random and take out as many coins as there were green marbles in the sample previously drawn, then place the unmarked envelope, with the remaining, unclaimed coins, in a large box. Because several participants performed the experiment at the same time we could not tell, for any individual participant, how many green marbles she had claimed to draw. At the same time, we could easily find out how many coins, on average, the participants had taken. Following that stage each participant engaged in the prediction task, in a different room.

Participants

Forty three students of the Hebrew University of Jerusalem were recruited for the experiment. The average age was 24.35 years ($SD = 2.43$ years). Nineteen of the participants were females.

Procedure

After the experimenter made sure that the participants could tell the difference between the marbles' different colors they entered the lab in groups of 4–8 participants in each session. Each participant worked alone at her own pace in a different cubicle. It was made clear to the participants that the experimenter could not see what they were doing. The participants got the materials – urn, tongs, instructions, pen, and a sheet of paper – and began the practice task, reporting into the computer. When they finished doing that, they raised their hand and the experimenter came in and gave them the written instructions for the second task. He made sure that they understood the instructions and went into the other room. The participants repeated the procedure, this time without the computer.

Participants were then instructed to select one of the unmarked envelopes and to take as many coins out of the 40 as the number of green marbles they had previously drawn from the urn. They then sealed the envelope and put it in the box.

After the participants put the envelope in the box, they took the same urn they used before and went to the next room, where they performed the prediction task as in Experiments 1 and 2.

Results

As mentioned before, the expected number of green marbles was 14. **Figure 5** presents the average number of green marbles as reported in the practice task, as derived from the number of coins removed from the envelopes, and predicted. A one-way ANOVA revealed a significant effect of stage [$F(2,126) = 4.66$, $MSE = 21.625$, $p = 0.011$, $\eta_p^2 = 0.069$). A *post hoc* comparison revealed that the mean number of coins claimed (corresponding to the values "reported" in the incentivized stage) was significantly different from both the value reported in the practice task and that predicted ($p = 0.031$, $p = 0.035$ respectively), whereas the values of the practice and the predicting task were not different from each other ($p = 0.999$). Furthermore, one sample t-test revealed that the average of the green marbles in the incentivized reporting task was significantly different from the expected value of 14 [$t(42) = 2.25$, $p = 0.016$], whereas the other two values were not [$t(42) = -0.751$, $p = 0.457$ for the practice task, and $t(42) = -0.599$, $p = 0.553$ for the predicting task]. The earnings in the prediction phase (Mean Payment = 0.837, $SD = 1.557$) were not significantly different from 1.139, that they would have

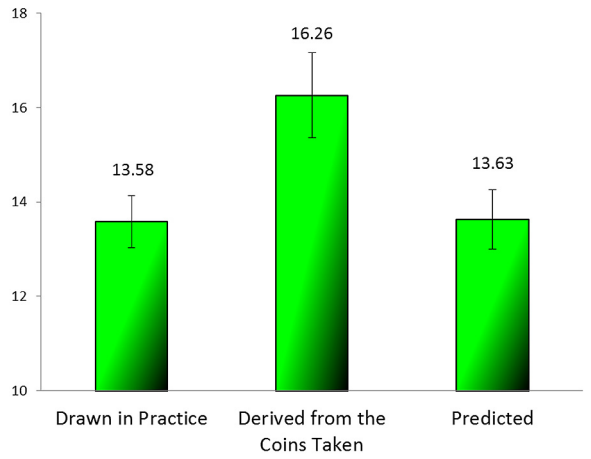

FIGURE 5 | Average of green marbles (those reported in the practice, those derived from the coins taken, and those predicted) (Experiment 3).

earned had they predicted 14 [$t(42) = -0.897$, $p = 0.375$) (see **Figure 6**).

The results of this experiment clearly demonstrate that when the lie was anonymous and there was no one who could call it out, the participants were no longer bound by it. Their predictions show that given such "non-bindingness" they easily made a prediction commensurate with the actual number of marbles of the infrequent color, and maximized their profits. It should be noted that, although in this experiment we couldn't distinguish between lying and steeling as did Mazar and Zhong (2010), given that participants were instructed to take out the number they sampled, the excess coins they took can be regarded as stealing by lying.

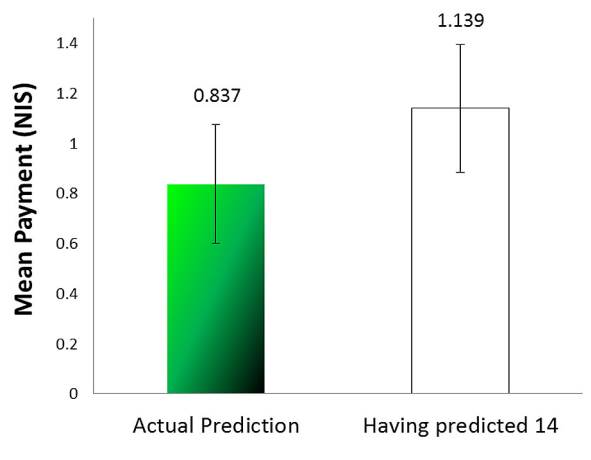

FIGURE 6 | Mean payment for prediction compared to a prediction of 14 (Experiment 3).

GENERAL DISCUSSION

The goal of this research was to study aspects of behavior following a lie. Specifically, we asked to what extent people would be ready to forgo a benefit in order not to imply, by a future action, that they had previously lied. We devised a sequence of two tasks, both involving the state of the same world. The first task allowed for profitable, voluntary lying behavior, in which the participants were assured, by the nature of the task that the experimenter could not tell if and when they had lied. Yet, a comparison of the overall statistical characteristics of a participant's report with the statistical characteristics of the environment could indicate the likelihood that lying had taken place. In the subsequent unexpected task, benefits would have been higher if the true state of the world, rather than that implied by one's previous reports, were used. As the second task was unexpected, the benefits of previously reporting the true state of the world could not be foreseen. This setup created a possible dilemma for liars, because deviation from their report in the initial task would constitute an implicit admission of having lied. The way our participants resolved the dilemma, when it existed, allowed us to assess the degree to which false reports bound the participants later on.

In Experiment 1, we have shown that people are willing to risk future profit or even forgo it altogether, in order not to get caught in a lie. The explanation we offer for such behavior is that people are "bound" to a lie they told, and are compelled by it to a certain behavior. That is, after providing the experimenter with a false report of the proportion of the profitable marbles, that person feels committed to that false representation in the sense that the person subsequently continues to predict a higher proportion than the proportion that would have most probably yielded a larger gain (but which would have been hard to justify in light of the previous report).

In Experiment 2, we have replicated these results and eliminated an alternative explanation based on anchoring, by repeating the reporting task with no incentive to lie – a task in which reporting turned out to be much closer to the value expected on statistical grounds.

In Experiment 3, we have shown that when there is no audience to the false presentation then there is no commitment to the lie. By creating a situation of a non-binding lie and showing that in that case participants felt free to act differently (and in line with what was more profitable) in the reporting and the predicting task, we have dismissed alternative explanations like an inability to correctly estimate the proportion of marbles of the infrequent color in the urn, self-deception, or anchoring.

It is important to note that the commitment to the lie does not stem from the risk of punishment, as even if inconsistent reports in the two tasks indirectly indicated that a person had lied, there was no sanctioning mechanism in our studies. Furthermore, when misrepresentation cannot be detected, as in Experiment 3, participants did not seek to be coherent with their reports.

The theory of self-concept maintenance (Mazar et al., 2008) explains well why participants did not lie "all the way" when they had an opportunity to do so. It is possible that the extra gains they made were exactly what struck a balance between monetary temptation and keeping one's self-image intact. This theory also provides an explanation for the self-deception in performance prediction described in Chance et al. (2011). In the latter, participants may have been unaware of how much they were aided by the answers sheet, and that enabled them to deceive themselves unlike in our experiments. In the experiments reported here every lie was very prominent for the participants as they held the marble of the unprofitable color in their hand and clicked the profitable color key on the computer screen (Experiments 1 and 2) or marked it on paper (Experiment 3). It might be that failing to maintain a self-concept of honesty, people proceed to (perhaps less desirable) honest impression management. It could be that when participants in Mazar and Hawkins (2015) who inflated their performance evaluation after lying, may also have engaged in impression-management rather than in deceiving themselves.

The contrast in prediction behavior between Experiment 3 and the previous 2 experiments shows that the commitment to the lie is of impression management and not a result of self-deception.

All in all our results indicate that people feel bound by their lies and that, once having told a lie, they are willing to risk future profits in order not to be exposed as having lied.

Not all lies commit liars. Additional research should address the motivation of lying and classify the situations in which lies bind the liar. We think that people sometimes use lies as means of impression management and, as long as possible, would prefer to deceive oneself rather than admit to have lied. But when this is impossible, *looking* honest would become an important target behavior, even if costly. Impression management could be particularly strong for lies in which one embellishes reality, presenting oneself as better as or more competent than one really is. At the same time, impression management using lies could have a positive effect through binding, by becoming a commitment to the lie, a motivation and a tool for establishing and improving self-identity.

REFERENCES

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychol. Bull.* 91, 3–26. doi: 10.1037/0033-2909.91.1.3

Baumeister, R. F., and Jones, E. E. (1978). When self-presentation is constrained by the target's knowledge: consistency and compensation. *J. Pers. Soc. Psychol.* 36, 608–618. doi: 10.1037/0022-3514.36.6.608

Camden, C., Motley, M. T., and Wilson, A. (1984). White lies in interpersonal communication: a taxonomy and preliminary investigation of social motivations. *West. J. Speech Commun.* 48, 309–325. doi: 10.1080/10570318409374167

Chance, Z., Norton, M. I., Gino, F., and Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proc. Natl. Acad. Sci. U.S.A.* 108 (Suppl. 3), 15655–15659. doi: 10.1073/pnas.1010658108

Cialdini, R. B., Trost, M. R., and Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *J. Pers. Soc. Psychol.* 69, 318–328. doi: 10.1037/0022-3514.69.2.318

DePaulo, B. M., and Kashy, D. A. (1998). Everyday lies in close and casual relationships. *J. Pers. Soc. Psychol.* 74, 63–79. doi: 10.1037/0022-3514.74.1.63

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Grice, H. P. (1989). *Studies in the Way of Words*. Cambride, MA: Harvard University Press.

Hample, D. (1980). Purposes and effects of lying. *South. Speech Commun. J.* 46, 33–47. doi: 10.1080/10417948009372474

Hilbig, B. E., and Hessler, C. M. (2013). What lies beneath: how the distance between truth and lie drives dishonesty. *J. Exp. Soc. Psychol.* 49, 263–266. doi: 10.1016/j.jesp.2012.11.010

Leary, M. R., and Kowalski, R. M. (1990). Impression management: a literature review and two-component model. *Psychol. Bull.* 107, 34–47. doi: 10.1037/0033-2909.107.1.34

Lewicki, R. J. (1983). "Lying and deception: a behavioral model," In *Negotiating in organizations*, eds M. H. Bazerman and R. J. Lewicki (Beverly Hills, CA: Sage), 68–90.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mazar, N., and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *J. Public Policy Mark.* 25, 117–126. doi: 10.1509/jppm.25.1.117

Mazar, N., and Hawkins, S. A. (2015). Choice architecture in conflicts of interest: defaults as physical and psychological barriers to (dis) honesty. *J. Exp. Soc. Psychol.* 59, 113–117. doi: 10.1016/j.jesp.2015.04.004

Mazar, N., and Zhong, C. B. (2010). Do green products make us better people? *Psychol. Sci.* 21, 494–498. doi: 10.1177/0956797610363538

Millar, K. U., and Tesser, A. (1988). Deceptive behavior in social relationships: a consequence of violated expectations. *J. Psychol.* 122, 263–273. doi: 10.1080/00223980.1988.9915514

Robinson, W. P., Shepherd, A., and Heywood, J. (1998). Truth, equivocation concealment, and lies in job applications and doctor-patient communication. *J. Lang. Soc. Psychol.* 17, 149–164. doi: 10.1177/0261927X980172001

Sagarin, B. J., Rhoads, K. V., and Cialdini, R. B. (1998). Deceiver's distrust: denigration as a consequence of undiscovered deception. *Pers. Soc. Psychol. Bull.* 24, 1167–1176. doi: 10.1177/01461672982411004

Schlenker, B. R. (1975). Self-presentation: managing the impression of consistency when reality interferes with self-enhancement. *J. Pers. Soc. Psychol.* 32, 1030–1037. doi: 10.1037/0022-3514.32.6.1030

Schurr, A., Ritov, I., Kareev, Y., and Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgm. Decis. Mak.* 7, 679–688.

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Human Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Silverman, L. J., Rivera, A. N., and Tedeschi, J. T. (1979). Transgression-compliance: Guilt, negative affect, or impression n management? *J. Soc. Psychol.* 108, 57–62. doi: 10.1080/00224545.1979.9711961

Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester: Wiley.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 August 2015
doi: 10.3389/fpsyg.2015.01075

The slow decay and quick revival of self-deception

Zoë Chance [1], Francesca Gino [2], Michael I. Norton [2] and Dan Ariely [3]*

[1] Yale School of Management, Yale University, New Haven, CT, USA, [2] Harvard Business School, Harvard University, Boston, MA, USA, [3] Fuqua School of Business, Duke University, Durham, NC, USA

People demonstrate an impressive ability to self-deceive, distorting misbehavior to reflect positively on themselves—for example, by cheating on a test and believing that their inflated performance reflects their true ability. But what happens to self-deception when self-deceivers must face reality, such as when taking another test on which they cannot cheat? We find that self-deception diminishes over time only when self-deceivers are repeatedly confronted with evidence of their true ability (Study 1); this learning, however, fails to make them less susceptible to future self-deception (Study 2).

Keywords: self-deception, cheating, self-enhancement, positive illusions, motivated reasoning

OPEN ACCESS

Edited by:
Eddy J. Davelaar, Birkbeck, University of London, UK

Reviewed by:
Shane Mueller, Michigan Technological University, USA
Gordon R. T. Wright, University of Leicester, UK

****Correspondence:***
Zoe Chance, Yale School of Management, Yale University, 165 Whitney Avenue, New Haven, CT 06511, USA
zoe.chance@yale.edu

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *11 November 2014*
Accepted: *13 July 2015*
Published: *19 August 2015*

Citation:
Chance Z, Gino F, Norton MI and Ariely D (2015) The slow decay and quick revival of self-deception. Front. Psychol. 6:1075. doi: 10.3389/fpsyg.2015.01075

Introduction

Imagine a stock trader who has access to insider information on particular firms, and as a result of using this information earns exceptionally high returns. If he then judges his stock trading ability by this performance, he may deceive himself into expecting high returns when he invests in other firms as well—discounting his cheating as the cause of his performance in favor of a self-deceptive view that the performance was due to his ability. Imagine that in his subsequent trades, he is lacking any insider information; over time, as a result, his future portfolio performance will give him unbiased evidence of his true ability. Will the trader eventually readjust his self-deceptive beliefs, and come to a more realistic understanding of his true ability?

We study both the decay and subsequent revival of self-deception in situations in which cheaters who have believed their superior performance was due to exceptional ability are then confronted with evidence of their true ability. How many doses of reality does it take before the truth sinks in and is accepted? After realizing the force and pitfalls of self-deception, are individuals less likely to engage in self deception in the future?

Motivated Views of the Self

People tend to see themselves through rose-tinted glasses. Decades of research document the tendency to self-enhance (Greenwald, 1980; Sedikides and Strube, 1997), with most people inflating their standing on positive attributes ranging from intelligence to ability to morality (Alicke, 1985; Taylor and Brown, 1988). Much of the empirical work on biased self-evaluations has explored the motivation for overestimating our own abilities or viewing ourselves as better than we truly are (e.g., Burson et al., 2006). This motivation is so strong that most people ignore or rationalize negative information about themselves to maintain a positive self-image (Pyszczynski and Greenberg, 1987; Kunda, 1990; Chance and Norton, 2010). They use motivated reasoning to interpret ambiguous information in ways that confirm their—generally positive—beliefs and attitudes about themselves (e.g., Lord et al., 1979; Ditto and Lopez, 1992; Swann et al., 1992). Moreover, people display impressive creativity in justifying questionable behavior and decisions (e.g., Norton et al., 2004; Gino and Ariely, 2012).

Self-deception

Although honesty is central to the self-concept (Blasi, 1984; Aquino and Reed, 2002), people routinely attempt to deceive others: in one diary study, participants reported lying once or twice a day (DePaulo et al., 1996). While some of these are "white" lies to protect others' feelings, many are self-serving. Rather than lying to maximize their economic utility, however, people often use a "fudge factor" that gives them some moral wiggle room—to lie or cheat just a little (Mazar et al., 2008). Farrington and Kidd (1977) show that people are more likely to dishonestly accept a smaller amount of money, and Goldstone and Chin (1993) show that people rarely fail to report making copies, but rather often underreport the number of photocopies they had made—even when they were not monitored.

Deceiving others has the potential benefit of getting ahead, even just to save a few pennies. But why would humans deceive themselves? Evolutionary psychologists have posited that self-deception evolved to assist in other-deception—the surest way to deceive others and not display signs of lying is to deceive oneself (e.g., Trivers, 2000). Most relevant to the present research, self-deception can allow people to hold preferred beliefs, regardless of the truth. Whereas motivated reasoning describes the general process of maintaining preferred beliefs, self-deception is a special case. "Stock examples of self-deception, both in popular thought and in the literature, feature people who falsely believe—in the face of strong evidence to the contrary—that their spouses are not having affairs, or that their children are not using illicit drugs, or that they themselves are not seriously ill" (Mele, 2001, p. 9). We follow Mele in defining self-deception as a positive belief about the self that persists in spite of disconfirming evidence.

Such beliefs can be maintained by attending to desirable evidence and avoiding conflicting undesirable evidence whenever possible. Greenwald (1997) compares knowledge avoidance to junk mail processing: if knowledge can be identified as unwelcome, a person may discard it before examining it thoroughly to learn precisely what it is. Self-deception is thus possible when ambiguity or vagueness leaves room for error or distortion (Gur and Sackeim, 1979; Baumeister, 1993; Mijovic-Prelec and Prelec, 2010; Sloman et al., 2010).

Chance et al. (2011) provided a new paradigm for demonstrating self-deception: participants who had an opportunity to cheat on a test by being given access to an answer key—and who therefore performed well—systematically overestimated their performance on future tests. Faced with the choice between attributing their performance to the presence of the answers or their own ability, people chose to self-deceive, convincing themselves that their performance was due not to the answers but to themselves. Importantly, Chance et al. (2011) incentivized participants for accurate predictions in one experiment. Whereas monetary incentives have eliminated face-saving lies in other studies (e.g., Dana et al., 2006), participants in the Chance et al. (2011) study who were paid for both performance and accuracy overpredicted their scores even when those overpredictions were costly—suggesting that overpredictions were self-deceptive rather than simply a no-consequence decision that allowed them to maintain consistency. As further evidence that the paradigm captures self-deception, overpredictions in the Chance et al. (2011) paradigm were correlated with trait self-deception, as measured by a scale of self-deceptive denial (Paulhus, 1998).

The Present Research

Previous experiments have examined self-deception as a momentary phenomenon. Life, however, offers many opportunities to act, to gather information, and to update beliefs—or not. In this work, we allow participants to cheat on an ability-based task to reap greater financial reward. We suggest that, rather than interpreting their behavior as a negative signal about themselves ("I'm a cheater"), self-deceivers use the positive outcome of cheating to bolster positive beliefs about themselves ("I'm a high achiever"). We add to the previous research on self-deception by using a modified version of the paradigm developed by Chance et al. (2011) to study whether and how quickly self-deception decays when individuals are confronted with repeated evidence of their actual ability. Building on the previous work, we also test how people's chronic tendencies to lie to themselves, and to others, relates to the pattern of overpredictions over time. Study 1 observes the decay of self-deception when an initial act of self-deception (inflating one's sense of one's abilities on the basis of a high score achieved by cheating) is followed by two rounds of unbiased feedback (scores on subsequent tests without an opportunity to cheat). Study 2 explores whether a second cheating opportunity can counteract the debiasing effect of feedback on actual abilities and reinstate self-deception. Together, these studies map the slow decay and quick revival of self-deception.

Study 1: The Decay of Self-deception

Study 1 examines the extent to which self-deception persists despite repeated evidence against a desired self-view. Participants completed a battery of four tests of general knowledge, predicting their score before the last three. Some participants—those in the answers condition—had access to an answer key for Test 1, and we expected them to use it to cheat (evidenced by outperforming a control group without answers). We also expected their high scores to trigger self-deception, leading them to overpredict their scores on subsequent tests for which they did not have answer keys. Performance on these subsequent tests offered repeated evidence of participants' true ability. We assessed the extent to which the inflated predictions of participants given the answers on Test 1 would be tempered by their later experience taking tests without the answers, hypothesizing that their predictions would eventually but not immediately converge with their true ability. Previous research has shown self-deception in this paradigm tracked with participants' chronic inclination to self-deceive (Chance et al., 2011), and we expected that the decay of self-deception here would be related to chronic self-deception as well. We hypothesized that for participants in the answers condition, self-deception would be greater and persist longer for those who were dispositionally high in self-deception. Furthermore, using an other-deception related scale in combination with the self-deception scale allowed us to test whether prediction gaps were indeed correlated with self-deception and not with lying.

Since the design of these self-deception studies makes cheating ambiguous—intentionally so, to make self-deception possible—we conducted a pilot study to test whether using the answer key did indeed constitute cheating. According to Jones (1991) definition of unethical behavior, community members, rather than researchers or participants given the opportunity to cheat, are the appropriate judges of which behaviors constitute cheating. Sixty-five participants from Amazon's Mechanical Turk read a description of our experimental research paradigm, including the instructions to participants, learned the results, and were asked to write four words describing the test takers. In their open-ended responses, "cheating" was the second most common open-ended response (15 people), after "dishonest" (22 people); 86% used the words "cheating," "dishonest," "unethical," or synonyms of these words. Participants also rated the extent to which they considered this behavior to constitute cheating, on a 10-point scale (1: *definitely not cheating* to 10: *definitely cheating*). The mean response was 6.98 (SD = 2.86), with a modal response of "10." Another group of 64 participants read about participants in the control condition, and indicated on the same scale whether that group was cheating; the mean response was 2.50 (SD = 2.63); the modal response was "1" (definitely not cheating). These results suggest that people judge the behavior of study participants in the answers condition who achieve higher scores to be unethical, such that "cheating" is an appropriate descriptor of their behavior. Cheaters do not need to perceive themselves as cheaters—indeed, they may be self-deceived.

Materials and Methods

Participants

Seventy-one student and community member participants (33 male, M_{age} = 23.9, SD = 3.54) from the paid subject pool of a large, northeastern university were paid $20 to complete this experiment as the first of a series of unrelated studies during a 1-h group lab session. Participants also had the opportunity to earn performance-based bonus pay. Sample size was determined by laboratory capacity, and privacy dividers separated participants from one another.

Design and Procedure

Each participant was assigned to either the control or the answer condition. Both groups completed a series of four tests of general knowledge trivia, such as "What is the only mammal that truly flies?" (Moore and Healy, 2008), configured into four 10-question tests. Participants learned at the beginning of the study that in all four tests, they would earn a $0.25 bonus for each correct answer. This incentive encourages cheating, which is required for self-deception in this paradigm, although a monetary incentive is not always necessary for prompting cheating and self-deception (Chance et al., 2011).

For Test 1, participants in the answers condition had the answers to all ten questions printed in an answer key at the bottom of the page. Their instructions read, "It's okay to check your answers as you go, but please do your own work." These instructions were intentionally ambiguous—they did not prohibit looking at the answers, but they did imply that using the answer key to choose answers would be wrong. The control group completed the same test questions but without the answer key or instructions. All participants were given 3 min to complete Test 1. After handing their completed Test 1 to an experimenter, they were given a score sheet with an answer key, on which they recorded from memory which questions they had answered correctly. This procedure prevented participants in the control group from using the answer key to change their answers. It did not prevent either group from inflating their reported score, therefore we recorded the actual score as well. After completing and turning in the score sheet, participants in both conditions had seen the answers for Test 1 and knew their Test 1 scores.

When participants received Test 2, they were asked to look it over before writing down their predicted score. The preview ensured that those in the answers group could confirm that the test would not include an answer key. It also reduced the implicit admission of guilt that might be associated with predicting a lower score on the second test than the first ("If I say I will do worse, the researchers will know I cheated"), by giving participants a valid excuse ("I just don't happen to know these particular answers"). Thus, this design provided a strong test of our prediction that participants who had cheated on the first test would deceive themselves into predicting an unrealistically high score on the second.

After predicting their score, participants spent 3 min completing Test 2, then repeated the process three more times: scoring Test 2 on a separate answer sheet; looking over Test 3 and making a prediction; scoring Test 3 on a separate answer sheet; looking over Test 4 and making a prediction; and scoring Test 4 on a separate answer sheet. Note that for all participants, Tests 2, 3, and 4 did not include answers at the bottom; and participants had only one sheet in front of them (either a test/prediction sheet or an answer key/score sheet) at all times.

When participants had finished the testing procedure, they moved on to other unrelated studies which also included the Balanced Inventory of Desirable Responding (Paulhus, 1998). We used the self-deceptive enhancement and the impression management components of the BIDR, to distinguish dispositional self-deception from dispositional lying. At the end of the study session, participants received their bonus payment. Because participants were not deceived (by the experimenters), the university Human Subjects Committee approving the experiment determined that debrief was not required.

Results and Discussion

Cheating

We predicted participants in the answers condition would inflate their performance on the first test by looking at the answers. Indeed, they reported scoring higher than the control group, $t(69) = 6.62$, $p < 0.001$, $d = 1.58$ (**Table 1**). Our subsequent analyses reflect reported scores, since self-deception relies on beliefs; however, using actual scores here or in any of the subsequent analyses did not affect the direction or significance of the results.

On the test in which cheating was possible, the average score was 7.89 out of 10, indicating either a mixture of cheaters and non-cheaters, many people cheating just a little, or both. Whereas

TABLE 1 | Study 1 scores and predictions.

		Test 1	Test 2	Test 3	Test 4
Answers	Prediction		6.28	5.72	5.53
	Score	7.89*	4.94	5.06	5.11
Control	Prediction		5.46	5.06	5.03
	Score	4.51	5	4.77	4.86

Answer key available, cheating possible.

no participants in the control condition reported perfect scores (a "10") on Test 1, 44% of participants in the answers condition did. However, even excluding perfect scores, Test 1 scores were higher in the answers than the control condition (6.20 vs. 4.51). This suggests many people cheating just a little, consistent with Mazar et al. (2008) theory of self-concept maintenance, which posits that people avoid negative self-signals by cheating only within an acceptable range.

Behavioral Self-deception

We expected that if participants in the answers condition were self-deceived, their predictions for subsequent tests would be higher than their actual scores; we expected this gap to be highest on Test 2—immediately after participants had cheated to achieve a high score on Test 1—and to decline over time. We did not expect participants in the control condition—who were not given the opportunity to cheat—to show a gap between their predictions and actual performance on Tests 2 through 4.

A paired t-test confirmed that Test 2 predictions exceeded Test 2 scores for participants in the answers condition, $t(35) = 3.67$, $p = 0.001$, $d = 0.73$ (**Table 1**) reflecting self-deception: despite having had the chance to examine the questions on Test 2 and confirm no answers were included, participants in the answers group expected to perform better than they did. Their surprisingly low scores on Test 2 did not eliminate their self-deception: their predictions for Test 3 were also significantly higher than their Test 3 scores, $t(35) = 2.52$, $p = 0.02$, $d = 0.35$ (**Table 1**). Only after scoring below their expectations on both Tests 2 and 3 did self-deception decay completely: predictions for Test 4 were not significantly higher than actual scores, $t(35) = 1.13$, $p = 0.27$, $d = 0.20$ (**Table 1**).

By contrast, predictions did not differ significantly from scores for participants in the control group for any of the three tests: Test 2 [$t(34) = 1.36$, $p = 0.18$], Test 3 [$t(34) = 0.95$, $p = 0.35$], Test 4 [$t(34) = 0.67$, $p = 0.51$] (**Table 1**). The lack of overprediction in the control group also indicates the inflated predictions of participants in the answers condition are not related to mere overconfidence: overconfidence would suggest people might generally inflate their predictions (Moore and Healy, 2008), but this pattern was not observed.

Dispositional Self-deception

We also explored whether the general tendency to self-deceive would relate to the decay in the observed prediction-performance gaps. Self-Deceptive Enhancement was indeed correlated with overpredictions on the second test ($r = 0.40$, $p = 0.02$) in the answers condition, but not the control condition ($p = 0.79$).

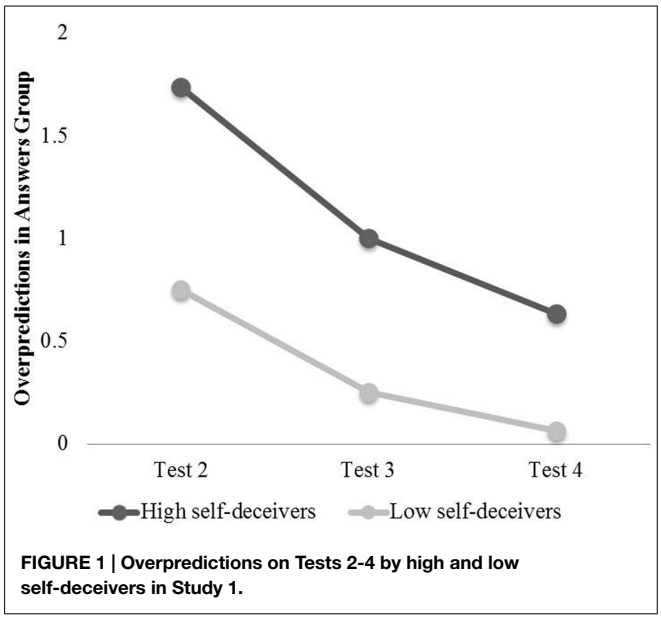

FIGURE 1 | Overpredictions on Tests 2-4 by high and low self-deceivers in Study 1.

A median split on Self-Deceptive Enhancement revealed that high self-enhancers were driving the self-deceptive predictions observed in the answers group, and that their bias was strong even in predictions for Test 3. High self-deceivers significantly overpredicted their scores on Test 2 [6.58 vs. 4.84, $t(18) = 3.07$, $p = 0.007$, $d = 0.93$] as well as Test 3 [5.95 vs. 4.95, $t(18) = 2.73$, $p = 0.01$, $d = 0.57$], but eventually even this group tempered their expectations to conform to reality, more accurately predicting their scores on Test 4 [5.74 vs. 5.11, $t(18) = 1.23$, $p = 0.24$]. Low self-deceivers in the answers group, on the other hand, did not show significant differences between any of their predictions and subsequent scores (all p's > 0.10). This pattern of results is shown in **Figure 1**. As expected, Impression Management showed no significant relationship to overpredictions in either the answers or the control group (all p's > 0.10), suggesting that the overpredicting observed here does not derive merely from a strategy to impress others such as the experimenters. For the answers group, we also compared Self-Deceptive Enhancement of those reporting perfect scores (likely cheaters) to those scoring lower; although the sample size was small and the observed difference not significant, those reporting perfect scores showed directionally higher Self-Deceptive Enhancement [7.19 vs. 6.20, $t(34) = 0.64$, $p = 0.52$]. Note that the self-deception observed here is not complete: participants in the answers condition do predict lower scores on Test 2 than they received on Test 1. These results suggest that rather than witnessing complete self-deception, we observe a self-deceptive miscalibration that then diminishes even more in the face of feedback.

These results demonstrate that self-deceivers come to terms with reality only when faced with repeated exposure to counterevidence against their preferred beliefs—for these participants, scoring lower on multiple tests they could not cheat on—and do so eventually rather than immediately. This pattern is most striking for those with a dispositional tendency toward self-enhancement.

Study 2: The Revival of Self-deception

Study 1 showed that when a single episode of cheating results in superior performance, it can lead to self-deception, but that repeated corrective feedback diminishes self-deception over time. However, in addition to providing evidence of a person's true abilities, life also offers repeated temptations to engage in questionable behavior, and thus repeated opportunities to self-deceive. Could later opportunities to cheat reinstate self-deception, overwhelming the educational effect of corrective feedback?

In Study 2, after some participants had cheated on Test 1 and had then taken Test 2 without an answer key and received legitimate feedback, we gave them a second chance to cheat by providing them with answers for Test 3. We predicted that those with the answer key for Test 3 would cheat again, and that their inflated scores would revive self-deception, evidenced by inflated predictions of their scores on Test 4.

Materials and Methods

Participants

One hundred forty-eight student and community member participants (68 male, Mage = 23.0, SD = 2.10) from the paid subject pool of a large, northeastern university were paid $20 to complete this experiment as the first of a series of unrelated studies during a 1-h lab session. Participants also had the opportunity to earn performance-based bonus pay. Sample size was determined by laboratory capacity, and privacy dividers separated participants from one another.

Design and Procedure

The design, procedure, and incentives in Study 2 were similar to those in Study 1. Briefly, participants took four tests and earned $0.25 for every correct answer. After each test was completed and scored, and after they had seen the answers, they looked over the next test, predicted their score, and completed the test. The only difference between Study 2 and Study 1 was that participants in the answers condition had an answer key at the bottom of Test 3 as well as Test 1.

Results and Discussion

Cheating

We predicted that participants in the answers condition would cheat when they had the opportunity, reporting higher scores than the control group. This was true in both cases in which they had the answer key, Test 1 [$t(146) = 8.07$, $p < 0.001$, $d = 1.33$] and Test 3 [$t(146) = 8.79$, $p < 0.001$, $d = 1.46$] (**Table 2**).

TABLE 2 | Study 2 scores and predictions.

		Test 1	Test 2	Test 3	Test 4
Answers	Prediction		6.06	6.95	5.75
	Score	7.65*	5.55	7.46*	5.28
Control	Prediction		4.97	4.51	4.79
	Score	5.03	5.34	4.68	4.88

Answer key available, cheating possible.

Self-deception

We also predicted, as in Study 1, that participants who had the opportunity to cheat on Test 1 would self-deceive: we expected their Test 2 predictions to be higher than their actual scores. A paired *t*-test confirmed that Test 2 predictions were indeed higher than Test 2 scores for participants in the answers condition [$t(79) = 2.69$, $p < 0.01$, $d = 0.24$] but not for those in the control condition, who predicted marginally lower scores than they achieved [$t(67) = 1.87$, $p = 0.07$] (**Table 2**).

When participants in the answers condition predicted their Test 3 scores, they did so with the knowledge of the answer key at the bottom of that test. We had no specific hypothesis regarding these predictions because we were interested in determining how cheating on Test 3 might influence their predictions for Test 4. We found Test 3 predictions for those in the answer key group were lower than the scores [$t(79) = 2.59$; $p = 0.01$, $d = 0.23$], whereas predictions for those in the control condition did not differ from the scores [$t(67) = 0.91$; $p = 0.37$] (**Table 2**).

Our key hypothesis in this study was that participants in the answers condition would reengage in self-deception after the second opportunity to cheat, and would predict unrealistically high scores on Test 4. As expected, they did so [$F(79) = 6.73$, $p = 0.01$, $d = 0.23$], whereas those in the control group did not predict unrealistically high scores [$F(67) = 0.12$, $p = 0.73$] (**Table 2**). A second opportunity to cheat appears to have reinstated self-deception, overcoming any learning from the unbiased feedback on Test 2.

General Discussion

One might expect people who cheat on tests—or insider traders—to feel worse about their abilities as a result of their questionable behavior. After all, if they had been more talented, they would have had no reason to cheat. However, when self-deception is possible, ethics can fade (Tenbrunsel and Messick, 2004). People tend to focus on the positive outcome of their cheating and neglect the unsavory process that led to it.

Although the construct of self-deception has a long history in psychology, the nature of the process by which self-deception takes place is still subject to debate (Audi, 1997; Mele, 2010; Bandura, 2011; McKay et al., 2011; von Hippel and Trivers, 2011). In these two studies, we showed that though self-deception does occur rapidly, there is some decay over time, suggesting that self-deception may provide temporary boosts to the self-concept but that these boosts may be relatively short-lived given corrective feedback from the environment (Study 1). Additionally, Study 2 demonstrates that sensitivity to feedback depends on the extent to which it enables self-deception; feedback bolstering motivated beliefs in superior abilities seems to be given more weight than feedback about actual abilities. As a result, it appears as though people are vulnerable to serial self-deception, awaiting opportunities to inflate their self-views and only grudgingly adjusting them downward. Study 1 demonstrates that inflated predictions of subsequent performance in the answers group correlate with general self-deceptive enhancement, and have suggested that these results suggest that participants engage in self-deceptive miscalibration. Future research might disambiguate

total self-deception from general miscalibration by comparing predictions of own scores to predictions of others' scores, allowing an assessment of whether people demonstrate self-deceptive miscalibration only when they are the focal actor, or whether even observing others induces miscalibration.

In our studies, we explored self-deception using a specific set of tasks similar to test situations in which students might have the opportunity to cheat. Although our focus was the impact of self-deception on people's beliefs about their future performance, self-deception in similar contexts might also affect subsequent behavior. It could, for example, lead students to spend less time preparing for future tests, thus reducing their learning as well as hampering their future performance. It might also increase the likelihood of cheating again, by allowing people to feel good about themselves and their abilities when they cheat (and then self-deceive). Future research is needed to examine these negative behavioral consequences of self-deception, not only in the context of academic cheating but also in the many situations in which people inflate their performance by cheating and then deceive themselves about why they did so well.

Acknowledgments

We thank Mika Chance and Mindi Rock for their patient and enthusiastic assistance with data collection on this project.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *J. Pers. Soc. Psychol.* 49, 1621–1630. doi: 10.1037/0022-3514.49.6.1621

Aquino, K., and Reed, A. (2002). The self-importance of moral identity. *J. Pers. Soc. Psychol.* 83, 1423–1440. doi: 10.1037/0022-3514.83.6.1423

Audi, R. (1997). Self-deception vs. self-caused deception. *Behav. Brain Sci.* 20, 104. doi: 10.1017/S0140525X97230037

Bandura, A. (2011). Self-deception: a paradox revisited. *Behav. Brain Sci.* 34, 16–17. doi: 10.1017/S0140525X10002499

Baumeister, R. F. (1993). "Lying to yourself: the enigma of self-deception," in *Lying and Deception in Everyday Life*, eds M. Lewis and C. Saarni (New York, NY: Guilford Press), 166–183.

Blasi, A. (1984). "Moral identity: its role in moral functioning," in *Morality, Moral Behavior, and Moral Development*, eds W. Kurtines and J. Gewirtz (New York: Wiley), 128–139.

Burson, K. A., Larrick, R. P., and Klayman, J. (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *J. Pers. Soc. Psychol.* 90, 60–77. doi: 10.1037/0022-3514.90.1.60

Chance, Z., and Norton, M. I. (2010). "I read playboy for the articles: justifying and rationalizing questionable preferences," in *The Interplay of Truth and Deception*, eds M. McGlone and M. Knapp (New York, NY: Routledge), 136–148..

Chance, Z., Norton, M. I., Gino, F., and Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15655–15659. doi: 10.1073/pnas.1010658108

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: costly (but quiet) exit in a dictator game. *Organ. Behav. Hum. Decis. Process.* 100, 193–201. doi: 10.1016/j.obhdp.2005.10.001

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

Ditto, P. H., and Lopez, D. F. (1992). Motivated skepticism: use of differential decision criteria for preferred and non-preferred conclusions. *J. Pers. Soc. Psychol.* 63, 568–584. doi: 10.1037/0022-3514.63.4.568

Farrington, D. P., and Kidd, R. F. (1977). Is financial dishonesty a rational decision? *Br. J. Soc. Clin. Psychol.* 16, 139–146. doi: 10.1111/j.2044-8260.1977.tb00209.x

Gino, F., and Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *J. Pers. Soc. Psychol.* 102, 445–459. doi: 10.1037/a0026406

Goldstone, R. L., and Chin, C. (1993). Dishonesty in self-report of copies made: moral relativity and the copy machine. *Basic Appl. Soc. Psychol.* 14, 19–32. doi: 10.1207/s15324834basp1401_2

Greenwald, A. G. (1997). "Self-knowledge and self-deception: further consideration," in *The Mythomanias: The Nature of Deception and Self-deception*, ed. M. S. Myslobodsky (Hillsdale, NJ: Erlbaum), 51–72.

Greenwald, A. G. (1980). The totalitarian ego: fabrication and revision of personal history. *Am. Psychol.* 35, 603–618. doi: 10.1037/0003-066X.35.7.603

Gur, R. C., and Sackeim, H. A. (1979). Self-deception: a concept in search of a phenomenon. *J. Pers. Soc. Psychol.* 37, 147–169. doi: 10.1037/0022-3514.37.2.147

Jones, T. M. (1991). Ethical decision making by individuals in organizations: an issue-contingent model. *Acad. Manage. Rev.* 16, 366–395.

Kunda, Z. (1990). The case for motivated reasoning. *Psychol. Bull.* 108, 480–498. doi: 10.1037/0033-2909.108.3.480

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *J. Pers. Soc. Psychol.* 37, 2098–2109. doi: 10.1037/0022-3514.37.11.2098

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–634. doi: 10.1509/jmkr.45.6.633

McKay, R., Mijović-Prelec, D., and Prelec, D. (2011). Protesting too much: self-deception and self-signaling. *Behav. Brain Sci.* 34, 34–35. doi: 10.1017/S0140525X10002608

Mele, A. (2001). *Self-deception Unmasked*. Princeton, NJ: Princeton University Press.

Mele, A. (2010). Approaching self-deception: how Robert Audi and I part company. *Conscious. Cogn.* 19, 745–750. doi: 10.1016/j.concog.2010.06.009

Mijovic-Prelec, D., and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 227–240. doi: 10.1098/rstb.2009.0218

Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychol. Rev.* 115, 502–517. doi: 10.1037/0033-295X.115.2.502

Norton, M. I., Vandello, J. A., and Darley, J. M. (2004). Casuistry and social category bias. *J. Pers. Soc. Psychol.* 87, 817–831. doi: 10.1037/0022-3514.87.6.817

Paulhus, D. L. (1998). *Manual for the Balanced Inventory of Desirable Responding*. Toronto: Multi-Health Systems.

Pyszczynski, T., and Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. *Psychol. Bull.* 102, 122–138. doi: 10.1037/0033-2909.102.1.122

Sedikides, C., and Strube, M. J. (1997). "Self-evaluation: to thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better," in *Advances in Experimental Social Psychology*, ed. M. P. Zanna (New York, NY: Academic Press), 209–269.

Sloman, S., Fernbach, P., and Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition* 115, 268–281. doi: 10.1016/j.cognition.2009.12.017

Swann, W. B., Jr., Stein-Seroussi, A., and Giesler, B. (1992). Why people self-verify. *J. Pers. Soc. Psychol.* 62, 392–401. doi: 10.1037/0022-3514.62.3.392

Taylor, S. E., and Brown, J. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210. doi: 10.1037/0033-2909.103.2.193

Tenbrunsel, A. E., and Messick, D. M. (2004). Ethical fading: the role of self deception in unethical behavior. *Soc. Justice Res.* 17, 223–236. doi: 10.1023/B:SORE.0000027411.35832.53

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Ann. N. Y. Acad. Sci.* 907, 114–131. doi: 10.1111/j.1749-6632.2000.tb06619.x

von Hippel, W., and Trivers, R. (2011). The evolution and psychology of self-deception. *Behav. Brain Sci.* 34, 1–56. doi: 10.1017/S0140525X10001354

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 24 December 2015
doi: 10.3389/fpsyg.2015.01965

When is Deceptive Message Production More Effortful than Truth-Telling? A Baker's Dozen of Moderators

Judee K. Burgoon *

Center for the Management of Information, Eller College of Management, University of Arizona, Tucson, AZ, USA

Deception is thought to be more effortful than telling the truth. Empirical evidence from many quarters supports this general proposition. However, there are many factors that qualify and even reverse this pattern. Guided by a communication perspective, I present a baker's dozen of moderators that may alter the degree of cognitive difficulty associated with producing deceptive messages. Among sender-related factors are memory processes, motivation, incentives, and consequences. Lying increases activation of a network of brain regions related to executive memory, suppression of unwanted behaviors, and task switching that is not observed with truth-telling. High motivation coupled with strong incentives or the risk of adverse consequences also prompts more cognitive exertion–for truth-tellers and deceivers alike–to appear credible, with associated effects on performance and message production effort, depending on the magnitude of effort, communicator skill, and experience. Factors related to message and communication context include discourse genre, type of prevarication, expected response length, communication medium, preparation, and recency of target event/issue. These factors can attenuate the degree of cognitive taxation on senders so that truth-telling and deceiving are similarly effortful. Factors related to the interpersonal relationship among interlocutors include whether sender and receiver are cooperative or adversarial and how well-acquainted they are with one another. A final consideration is whether the unit of analysis is the utterance, turn at talk, episode, entire interaction, or series of interactions. Taking these factors into account should produce a more nuanced answer to the question of when deception is more difficult than truth-telling.

Keywords: deception, cognitive effort, truth, deceptive message production, moderators of deception displays

OPEN ACCESS

Edited by:
Dan Ariely,
Duke University, USA

Reviewed by:
Giorgio Ganis,
Plymouth University, UK
Shaul Shalvi,
Ben-Gurion University of the Negev, Israel

***Correspondence:**
Judee K. Burgoon
judee@email.arizona.edu

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *03 September 2015*
Accepted: *07 December 2015*
Published: *24 December 2015*

Citation:
Burgoon JK (2015) When is Deceptive Message Production More Effortful than Truth-Telling? A Baker's Dozen of Moderators. Front. Psychol. 6:1965. doi: 10.3389/fpsyg.2015.01965

Common sense tells us that lying should be more difficult than truth-telling. After all, the truth is ready-made; the lie must be invented. Ceteris paribus, more effort is involved in fabricating a falsehood than in accessing and producing a veridical account of something that is already stored in memory.

But common sense is not always the best teacher. There are many circumstances under which truth-telling imposes more challenges than deceiving. I therefore want to advance the hypothesis that the effort associated with deceiving vice truth-telling is a function of the characteristics of the communication event in force and that deeper analysis of critical elements of the communication process will bring more clarity to the issue of the cognitive effort associated with

deceit. Although many such elements have been included as moderators in deception meta-analyses, their impact has not necessarily been attributed to cognitive (or emotional) exertion, and reliable empirical associations are few. A more coherent framework is therefore wanting.

THE DOMINANT PATTERN

First let us consider the received wisdom that deception is more difficult than truth and some of the evidence that undergirds it. Numerous deception scholars have argued that deception is more effortful than truth-telling (e.g., Zuckerman et al., 1981; Miller and Stiff, 1993; Buller and Burgoon, 1996b; Vrij, 2000; Sporer and Schwandt, 2006). Empirical research has affirmed this view with evidence of measurable psycho-physiological indicators of arousal and stress (e.g., the wealth of research on the polygraph; see Gougler et al., 2011) as well as observable behavioral signs of performance decrements. Deceptive messages are often shorter, slower, and less fluent, with longer response latencies, averted gaze, temporary cessation of gestures and postural rigidity–all potential indicators of deceivers having to think hard (Goldman-Eisler, 1958; Vrij et al., 1996, 2006; Rockwell et al., 1997; Porter and ten Brinke, 2010; ten Brinke and Porter, 2012; Mullin et al., 2014).

That said, it is important to note that the mental machinations associated with deception need not be burdensome or uniformly so. As Buller and Burgoon (1996b) stated in a rejoinder to DePaulo et al. (1996):

> …DePaulo et al. (1996) ascribe to us a highly cognitive view of deception, with deceptive episodes peopled by highly conscious, surveillant liars and equally vigilant, cunning receivers. This is an exaggerated characterization of our assumptions. We have taken some pains in IDT to argue that much sender and receiver activity during deceptive encounters, like other communicative encounters, can be goal driven and strategic yet largely automatic and "mindless" (see, e.g., Kellermann, 1992; Burgoon and Langer, 1995). We see deception running the gamut from the kinds of inconsequential white lies and evasions that populate daily discourse to the life-threatening kinds of fabrications and omissions that color international conflicts (Burgoon and Buller, 1996, pp. 320–321).

The activities involved in message production are familiar, routinized, overlearned. Mental processes can be activated without the sender necessarily having significant attentional resources diverted. This is especially likely in the dominant laboratory research paradigms, which entail telling harmless and inconsequential lies seldom lasting more than 1 min and addressing single incidents, factual matters, or likes-dislikes. In such cases, messages can be constructed on the fly and modified in response to emergent exigencies. Senders can tap into a host of memories and readily accessible schemas that enable rattling off a deceptive response. The division of labor between verbal and non-verbal components of messages further distributes the workload and reduces the call on cognitive resources. Moreover, if lies are about inconsequential matters, are at the behest of an investigator, and entail no adverse consequences, then any emotional overlay should also be attenuated.

That many forms of deception are "ready-made" does not invalidate that the other processes surrounding their use, form and potential consequences still impose more cognitive work on the sender than does a truthful message related to the same narrative. But the depiction of deceptive message production requires more sophisticated modeling. It is not a question of deception being *either* easier or more difficult than telling the truth. It can be both.

A BAKER'S DOZEN OF MODERATORS

Here, then, toward a more nuanced, communication-oriented view, are a baker's dozen of factors that should tip the scales in one direction or another. This non-exhaustive collection includes sender factors (i.e., ones that reside within the individual producing a message), message and communication context factors (i.e., ones related to the content and style of the message and to the communication context), relationship factors (i.e., ones inhering in the interpersonal relationship between sender and receiver) that should enable predictions of the circumstances under which deception will be more effortful, and scale of the measurement window under analysis. I illustrate many with evidence from our research program on interpersonal and mediated deception.

Sender Memory Demands

Recent neuroscience research is corroborating what social scientists have suspected for a long time—that the more a lie activates different mental processes, the more mental taxation it imposes on a communicator. In their updated conceptualization of cognitive resource demands associated with (complex) lie production, Sporer and Schwandt (2007) incorporated newer models of working memory such that cognitive load extends beyond accessing details from memory and constructing non-contradictory messages to also activating autobiographical and executive memory functions.

Consider that compared to the truth-teller, who needs only to recall an actual state of affairs, the deceiver must not only access the true state of affairs but must engage executive memory to decide *if* to deceive, evaluate which forms of deception are more "acceptable" according to one's moral code and choose among those options, conduct a cost-benefit calculus of the relative likelihood of success of alternative forms of deceit, fabricate the response itself, compare it to the truth for possible inconsistencies with known facts, check the deceit against a "plausibility" meter, gage the likelihood of suspicion or detection by the interlocutor, and then actually assemble the verbal and non-verbal components into a normal-appearing message that maximizes credibility, all the while suppressing inapt behaviors and cognitions.

Early explorations of brain functioning with fMRI confirmed that these activities have associated changes in brain activation such that different regions show increased activation during lies than truths (see, e.g., Spence et al., 2001; Ganis et al., 2003;

Abe and Greene, 2014). In one such test, Spence et al. (2008) found that the ventrolateral prefrontal cortex (VLPFC) was preferentially activated to inhibit inappropriate and unwanted cognitions and responses when lying about embarrassing material. Using a different method, Mameli et al. (2010) found multiple networks in the prefrontal cortex involved in deceptive responding as well as longer reaction times when communicators responded deceptively relative to truthful responses at baseline. Ito et al. (2011, p. 126) similarly substantiated increased activity in a network of brain regions in the dorsolateral prefrontal cortex (plus longer response latencies) when remembering and reporting truthful and deceptive neutral and emotional events. The authors did not find a similar response during truth-telling, leading them to suggest that "there is an increase in the amount of conflict and higher cognitive control needed when falsifying the responses compared to responding truthfully."

A recent meta-analysis (Christ et al., 2009) further established that lying is associated with multiple executive control processes, specifically working memory, inhibitory control, and task switching (i.e., interspersing truthful with deceptive details). Using their activation likelihood estimate method, the authors demonstrated quantitatively that eight of 13 regions and 173 deception-related foci are consistently more active for deceptive responses than for truthful ones.

These robust findings using varied approaches are strong evidence that deception summons memory processes that are more taxing than those associated with truth-telling. Thus, for the predominant research paradigms that have been used, and holding all other conditions constant, deception requires engagement of more cognitive (and/or emotional) resources than does truth-telling[1].

Sender Motivation, Incentives, and Consequences

This general pattern notwithstanding, three interrelated moderators that can alter this conclusion are motivation, incentives and consequences. Because motivation has often been manipulated through high monetary incentives or escaping adverse consequences, these three factors are operationally confounded. High motivation is thought to muster more effort, which can interfere with performance *or* improve it. The motivation impairment effect (MIE) asserts that motivation impairs non-verbal performance, thereby making lies more transparent, but also facilitates deceivers' verbal performance (DePaulo and Kirkendol, 1989; Bond and DePaulo, 2006). Empirical findings have been fraught with inconsistencies. Burgoon and Floyd (2000), Burgoon et al. (2012), and Burgoon et al. (2015) have found both impairment and improvement of non-verbal and verbal performance among motivated deceivers engaged in consequential deception. Additionally, high-motivation *truth-tellers* (not deceivers) sometimes were most affected. Two meta-analyses (that omitted the aforementioned investigations) found high motivation affected liars and truth-tellers equally (Bond and DePaulo, 2006), and high-motivation lies were neither more nor less detectable than other lies (Hartwig and Bond, 2014).

If communicators have little to gain from deceiving or to lose from being caught, lying may pose little more challenge than truth-telling. Aside from the memory demands discussed above, small everyday lies such as fibs and white lies are easy to produce, can draw upon a cache of previously used utterances, and countenance no danger if detected. Lies that are likely to summon more cognitive resources are those that yield high pay-off if successful or that place the deceiver in serious jeopardy if uncovered (Porter and ten Brinke, 2010). In an analysis of real high-stakes deception, ten Brinke and Porter (2012) found that deceivers feigning distress over their missing children had difficulty faking sadness, leaked expressions of happiness, and were verbally more reticent and tentative. The authors ascribed these performance decrements partly to increased cognitive load. In high-consequence circumstances, however, truthful individuals may be equally distressed or motivated to succeed, so the difficulty of producing believable messages may be similar regardless of veracity.

The diverse results suggest that motivation is more complicated than presupposed and requires more "unpacking" of its relationship to cognitive effort. From a communication standpoint, motivation should follow social facilitation predictions, aiding overlearned behavior and interfering with less practiced behavior, up to a point beyond which emotional flooding should impair both verbal and non-verbal performance. Communicator skill and experience should dictate the threshold for performance deterioration.

Discourse Genre

Language can be categorized according to genres, which are discourse forms that share similarities in their structure, style, content, intended audience, and context in which they occur. Different genres impose qualitatively different demands on deceivers and truth-tellers. A factual narrative or description, for example, comprises representational and verifiable features that need to be assembled into a cogent, plausible sequence, and supported by relevant details. Whereas truth-tellers are only limited by the acuity of their memory when relaying specifics of an event, deceivers not only must recall the true state of affairs, but must decide how much, if any, to tell. They must compare their alternative version to reality, edit the content and linguistic form, and assemble the elements into a believable chronology.

Comparatively, an opinion lacks verifiability and need not be accompanied by any supportive documentation. Deceivers can easily proffer indisputable conjectures and opinions when asked questions such as, "Who do you think may have stolen the money from the cash draw?" or "What should happen to the thief?", whereas the thoughtful reflections of a truth-teller may require more effort.

Within interactive discourse genres are also variations in form. A face-to-face dialog carries different demands than a monolog or one-to-many speech. When engaged in conversation with another, interlocutors must fulfill multiple communication

[1]Space limitations do not permit developing the idea that deception may also instigate emotional work to regulate the kind of emotional flooding seen, for example, with escalating conflicts. But investigations of high levels of cognitive arousal may be well consider emotional correlates and regulatory overrides.

functions beyond message production itself. First, they must "read" the definition of the situation from contextual cues so as to know what kind of discourse and associated expectations are in force. Because ascertaining identities is usually a high priority, communicators must signal their self-identity (e.g., gender, ethnicity, race, personality), put forth a desired self-presentation, and size up others' identities. As interactions unfold, they must formulate their own messages and decipher the messages and feedback from their interlocutor. They must also regulate their emotional expressions, exchange relational messages that define the relationship between sender and receiver (e.g., trusting, intimate, equal), perform turn-taking responsibilities, and monitor their own communication. Although human communicators perform these functions in a seemingly effortless fashion, the discourse form can magnify or alleviate some of the effort associated with them. For example, Burgoon et al. (2001) demonstrated that engaging in dialog compared to face-to-face monolog was more difficult initially, but over time, dialog eased the demands on deceivers who were able to share the turn-taking burden with their interlocutor, create a smooth interaction pattern by developing interactional synchrony, adapt to interlocutor feedback, and approximate normal communication patterns[2].

Another genre, the interview, can also influence the cognitive burden on respondents. The question-answer structure adds predictability to who is supposed to talk when and what the content should be. Language can be borrowed from the interviewer's questions, and questions can be repeated as a stalling technique. Even within interviews are notable differences: Relative to an open-ended, free-wheeling interview, a structured one that requires short-answer replies reduces the degrees of freedom of what can be said and allows deceivers to forecast what is coming next. Many deception experiments are of this latter brief-answer variety, which our research has shown produces substantially different behavioral and psycho-physiological responses than open-ended interview protocols (Burgoon et al., 2010).

The illustrative genres mentioned here point to the need to formulate deception-relevant taxonomies of genres so that predictions can be made as to which will intensify or diminish the cognitive effort required of sender and receiver.

Form of Prevarication

Contrary to the claims of McCornack et al. (2014) that virtually all extant deception research bifurcates deception into bald-faced lies or bald-faced truths, and regards only those discourse options as worthy of scholarly investigation, most deception scholars recognize that deception includes a variety of forms. A sampling of research across the last five decades and across multiple disciplines has identified such forms of prevarication as white lies, altruistic lies, omissions, concealment, equivocation, evasions, exaggerations, strategic ambiguity, and impostership (see, e.g., Turner et al., 1975; Hopper and Bell, 1984; Miller and Stiff, 1993; Buller et al., 1994; Searcy and Nowicki, 2005; Ennis et al., 2008; Knapp, 2008). The type of prevarication being told will affect the cognitive resources required in its telling.

In his original formulation of information manipulation theory (IMT), McCornack (1997) proposed that deceptive discourse violates conversational implicatures along one or more of Grice's (1989) four dimensions of cooperative discourse: quantity, quality, manner, and relation. Burgoon et al. (1996) proposed a similar set of five dimensions of information management: completeness (comparable to quantity), veridicality (comparable to quality), clarity (comparable to manner), relevance (comparable to relation), and personalism (see also Buller and Burgoon, 1996a). Under both conceptualizations, some forms of deceit such as omissions are more easily produced than others[3].

Other times, truth-telling can be more difficult than deceit. Having to convey a "hard" truth to a patient dying of a terminal disease can levy more cognitive taxation than manufacturing a comparable falsehood that there is hope for recovery from the disease. A provocative line of research on whether people lie automatically or must decide to lie has also shown that when cheating offers a high probability of personal gain, people may be quicker to produce self-serving lies than truthful responses. In tempting situations, if a self-benefiting lie is easy to craft and little time is allowed for reflection, lying may be the more automatic response, whereas honesty may necessitate more hesitation, deliberation, and executive control (Shalvi et al., 2012; Tabatabaeian et al., 2015; see also Bereby-Meyer and Shalvi, 2015, for a review of supporting literature). When social bonds are made salient, people also produce lies more quickly that benefit their social group than lies that benefit only self (Shalvi and De Dreu, 2014).

In short, the type of prevarication (or truth) can be located on a continuum from easy to difficult, with cognitive effort for easy lies making them no more challenging than telling the truth.

Expected Response Length

Different kinds of interactions have associated expectations about utterance length. Day-to-day conversations are typified by reciprocation of short turns at talk. Conversing deceivers may project that they can get away with very brief responses while still satisfying conversational expectations. A spouse's query, "How was your day?" is not expected to produce a dissertation on all

[2]Although some meta-analyses have attempted to analyze the effects of communication context or genre on receiver detection accuracy (e.g., Bond and DePaulo, 2006; Hartwig and Bond, 2014), virtually no research has explicitly tested their effects on sender performance. Hartwig and Bond (2014), for example, had too few samples of different interview types to separate out different categories. Part of the challenge in deriving stable meta-analytic estimates is that only a small fraction of investigations have entailed interactions exceeding 1 min in length. Moreover, genre constructs such as interactivity are multidimensional. To test properly the effects of interaction on senders requires parsing the different attributes (e.g., participation, synchronicity, propinquity, multiplicity of modalities) and testing each independently to isolate the relevant features.

[3]The least taxing form is concealment or omission in which deceivers simply leave out deceptive information. Although McCornack et al. (2014, p. 353) assert in IMT2 that "Zipf's PLE [principle of least effort] compels speakers to minimize the total number of spoken words produced and shift instead toward objectively ambiguous language," a claim consistent with the principle that humans are cognitively lazy, it fails to comport with the empirical evidence that people sometimes produce longer messages when deceiving than when telling the truth (e.g., Burgoon et al., 2014; Dunbar et al., 2014). Brevity, then, or effort is not the controlling factor.

one's trials and tribulations at work or home. A husband who skipped work to go gambling or a wife on an illicit tryst can safely reply with a breezy "fine." Such brief lies and truths—the bread and butter of much deception research–may differ little in their demands on resources. More penetrating questions like, "Why couldn't I reach you today when I called your cell four times?" require lengthier–and more demanding–accounts.

Standard interview protocols also have associated expectations about what response lengths suffice. Introspective questions require conjectural rather than factual responses, and their non-verifiability may attenuate the memory burden on deceivers. The behavioral analysis interview operates on the premise that innocent people will exhibit the Sherlock Holmes effect: In attempting to aid an investigation, innocent respondents may speculate more than deceivers and widen the pool of suspects. Comparatively, deceivers should minimize conjecture and avoid proposing other suspects for fear of narrowing the pool to themselves (Horvath et al., 2008). A cognitive interview, in which respondents are asked to retell an account from multiple vantage points (Fisher and Geiselman, 1992), requests increasing elaboration and details, something that is expected to be easier for truth-tellers than deceivers to accomplish over repeated retellings (see also Vrij and Granhag, 2012).

Generally, conversations have associated norms and expectations for what kinds of utterances will satisfy the Gricean maxims, and communicators are fairly adept at predicting and fulfilling those expectations. The degree of cognitive difficulty should correlate positively with response length and how much the deceptive response deviates from expected form (with exceptions that can be anticipated in advance).

Sanctioning of Deceit

Most laboratory research involves deceit that is sanctioned by the experimenter rather than being chosen voluntarily by the perpetrator (Frank and Feeley, 2003). The alternative of allowing research participants to choose whether to lie or not creates a confound in that only skillful liars and those with an honest-appearing demeanor may choose to lie (Levine et al., 2010). Apart from experimenter-instigated deceit differing behaviorally from that chosen of a deceiver's own volition (Sporer and Schwandt, 2007; Dunbar et al., 2013), the implication outside the laboratory is that deception will vary substantially in form and difficulty as a function of sanctioning and communicator skill (see also IDT regarding communicator skill).

That said, choice and skill may not completely alleviate the added cognitive work associated with deceit. Spence et al. (2008) designed an fMRI experiment in which deceivers could choose to comply or defy an experimenter's request to divulge embarrassing secrets. Results revealed lying activated the VLPFC even under free choice. At the most fundamental level of brain functioning, then, lying still exercises a main effect on cognitive processing.

Communication Medium

The medium of communication itself also influences the degree of cognitive difficulty associated with lying. IDT's first proposition states, "*Context features of deceptive interchanges systematically affect sender and receiver cognitions and behaviors; two of special importance are the interactivity of the communication medium and the demands of the conversational task*" (Burgoon and Buller, 2015). To the extent that deceivers are interacting synchronously and with all audiovisual modalities available to receivers (e.g., face-to-face, computer-mediated communication, teleconferencing), there are more communication functions to which cognitive resources must be devoted. When modalities are more limited–such as voice or chat–and asynchronous—more resources can be distributed among fewer aspects of message production and with less time press.[4] Consistent with this reasoning, participants in a mock theft experienced the least anxiety and cognitive load when interacting via text, were the most aroused and exercised the most behavioral control when interacting face-to-face, and reported the most cognitive effort when interacting via an unfamiliar audio format (Burgoon et al., 2004; Burgoon, 2015). Thus, leaner and non-interactive media should attenuate cognitive effort.

Preparation

This construct subsumes many related variables—advance thought, planning, rehearsal, or editing. Extemporaneous or unscripted discourse is produced in real time; planned, rehearsed, or edited discourse entails some intervening time interval between the deliberation and construction of a message and its ultimate delivery. Such *ex ante* preparation may be experimentally manipulated, as in a classic interviewing investigation by O'Hair et al. (1981), or it may be prompted by high-stakes circumstances such as queries about fraudulent financial reporting: "…individuals may, for example, prepare extensively before speaking to lower the cognitive burden that can accompany deception, or may undergo voice training in an attempt to sound vocally like the antithesis of someone engaging in deception" (Burgoon et al., 2015, p. 2).

Three meta-analyses (Zuckerman and Driver, 1985; DePaulo et al., 2003; Sporer and Schwandt, 2006) included preparation as a moderator and predicted that planning and rehearsal should facilitate deceptive performance by reducing cognitive/memory load. Although the meta-analyses yielded mixed results and weak effect sizes, planned messages were found to have shorter responses latencies and fewer silent pauses than unplanned ones. More recent research examining higher stakes deception has shown that fraud-relevant utterances were longer and more laden with details than non-fraudulent ones (Burgoon et al., 2015), a pattern duplicated by Braun et al. (2015) in their analysis of deceptive politicians' messages. To the extent that detection accuracy is lower with planned than unplanned deception (Bond and DePaulo, 2006), some of that inaccuracy may be attributable to planned messages being indistinguishable from truth-telling. With advance preparation, communicators are better able to approximate normal, credible communication patterns.

[4]It might be tempting to conclude that we can infer the degree of cognitive demands on senders by the accuracy with which their messages are detected by receivers. However, this would be a faulty inference inasmuch as detection accuracy is influenced by several factors other than sender performance (Burgoon et al., 2008; Burgoon, 2015). For example, deceivers may experience fewer cognitive demands under audio communication and yet inadvertently produce more telltale signs of deception due to lack of awareness or ability to manage the voice.

Recency of Target Incident or Issue

Depending on how distant it is, the time frame for requested narratives and accounts will have expectations associated with it for what is a complete, accurate, and clear response. Whereas recent events should impose equal recall difficulty on truth-tellers and deceivers, long-ago ones should be harder to recall for conscientious truth-tellers trying to be thorough and accurate than for deceivers fabricating a story or borrowing details from similar events. Some interview protocols like the cognitive interview capitalize on this reversal of expectations in which longer and more effortful answers should be associated with truth. Comparison questions in polygraph testing which are intended to create more mental conflict for truth-tellers than deceivers can be made even more challenging when the time frame is open-ended. The question, "Have you ever lied to someone who trusted you?" may prompt truth-tellers to ponder and hesitate more than deceivers. Other aspects of cognitive work unique to deceivers are the activation of executive memory to make the decision to lie, the construction and selection among possible lies and the comparison to the truth, which may guide decisions about which form and content of the lie is likely to be the most efficacious.

Cooperative-Adversarial Relationship

Intertwined with the genre of discourse is whether the relationship between communicators constitutes a cooperative or adversarial one. Grice (1989) proposed that communicators enter encounters with a presumption of cooperativeness. In practice, however, many communication contexts and relationships are recognized as adversarial–criminal interrogations, litigation, labor disputes, negotiations, dispute mediations, and divorce proceedings that place the parties at odds with one another, among others–during which the assumption of cooperativeness is suspended. In adversarial interactions, one cannot even assume that interlocutors are using language in the same way. For example, in organizational contexts, management may practice strategic ambiguity as a way to reduce rather than facilitate understanding.

In other cases, participants with hidden agendas may wish to give the appearance of cooperativeness while covertly violating the Gricean maxims (McCornack, 1997). Under these circumstances the success of the deception will depend on how clandestine the deceit is. Predictions about how much cognitive difficulty is associated with lying should take into account how much cognitive "work" is needed to keep nefarious motives hidden. Unwitting interlocutors, for example, may lessen the difficulty for deceivers by proposing plausible explanations for a sender's otherwise implausible response, thereby helping deceivers construct a believable narrative as a dialog unfolds.

Relational Familiarity

Buller and Burgoon (1996b) identified three types of familiarity, one of which is relational familiarity. People who are well acquainted with one another have prior knowledge and a history of behavior against which to judge anything that is said. For the deceiver, this can make devising a plausible lie that evades detection more challenging inasmuch as there are numerous touchpoints against which the deceiver must make mental comparisons before actually uttering the lie. At the same time, deceivers can capitalize on their familiarity with the receiver to adapt lies more specifically to the interlocutor's knowledge bank and can watch the receiver for telltale signs of disbelief. Buller and Aune (1987) found deceivers interacting with familiar others successfully restored their original level of animation, while deceivers interacting with strangers became less immediate and animated over time. Thus, deceivers took advantage of their relationship to improve their performance over time. Burgoon et al. (2001) found similar results in that deceivers interacting with friends rather than strangers were better able over time to manage their informational content, speech fluency, non-verbal demeanor, and image. Presumably the improved performances were accompanied by a corresponding reduction in cognitive difficulty for deceivers relative to truth-tellers. Since receivers seldom expect to be lied to, relational familiarity probably confers more of an advantage on the sender than the receiver.

Communication Unit of Analysis

The sampling unit for deception research and meta-analyses typically has been the single utterance, turn at talk, or answer to a single question. Such samples may be less than 30 s in length. Yet deception may be woven into a series of utterances (e.g., an interview), interpenetrate an entire conversational episode, or span multiple conversations (e.g., multiple interrogations). The span of time from beginning to end of a deception event should affect how difficult it is to produce and maintain. Speculatively, as the number and duration of utterances related to an issue increases, the more cognitively challenging it should be to lie, inasmuch as one must remember what has been said previously, create consistency among utterances, reconcile what is being said with a potentially growing population of known facts, make decisions about which truthful details to divulge, decide what kinds of deception to enact, whether to change strategies (e.g., from concealment to equivocation), and so forth. Lengthy criminal justice interviews and interrogations depend on extended questioning to create more emotional and mental hardship for interviewees. Comparatively, producing brief utterances not only minimizes the amount of decision making, memory searching and message production demands that communicators incur (regardless of their veracity) but can also buy deceivers more time to concoct a credible response and to intersperse truthful details within one's discourse to bolster believability.

The time course of the communication event thus may dictate its demand on cognitive and emotional resources. As the number of utterances or interchanges increases, demands on cognitive and emotional resources should increase differentially—up to an as-yet undetermined point. Beyond that, cooperative interactions should reduce the burden on deceivers by virtue of availing themselves of receiver feedback, making conversational repairs and meshing the dyad's interaction patterns. We have witnessed this in several of our interviewing experiments. In one case, interviewees who were blindsided by unexpected

questions initially gave non-fluent and improbable responses but with the aid of unwitting interviewers managed to spin out explanations that the interviewers accepted. Conversely, adversarial interactions such as interrogations may intensify the burden on deceivers. In drawing any conclusions, then, about whether lying is more difficult than truth-telling, it is necessary to specify the sampling unit for the respective truths and lies—short utterances or lengthy ones and single episodes or a series of them. Longer can be more difficult but may also introduce opportunities for countervailing repairs by deceivers.

IMPLICATIONS

What are the implications of this decomposition of moderators of cognitive effort? First, the relationship between deception and cognitive effort is complex and highly variable. In some respects, the issue is one of definition of terms: What constitutes effort? If activation of more brain regions and processes constitutes effort, then deceit can be construed as creating greater *actual* cognitive work than truth. However, if effort requires some level of awareness, then only under more serious circumstances involving complex lies with significant (favorable or unfavorable) consequences may lying be *experienced* as more cognitively effortful.

Moreover, a variety of moderators can alter the deception-cognition relationship, and sometimes in contradictory ways. These previously unidentified or untested moderators may account for the oft-times weak association between presumed cognitive effort and observable behavior. Only if the relevant influences can be parsed will it be possible to make sound and reliable cognition-based predictions and will cognition-based effects be replicable.

Also confounding the picture is that many factors like motivation and incentives exert similar influence on truth-tellers, thus making deceptive and truthful behavior patterns indistinguishable.

Too often, researchers have inferred backward from observable cues to likely cognitive causes, but such reasoning is fraught with indeterminacy due to the absence of single one-to-one correspondences between specific indicators and mental work. Even though more memory processes may be engaged, the observable indicators may not betray that work, they may arise from other causes, and they may be associated with both truth and deception.

Given these complicating factors, any cognitive load, cue-based approach may be difficult to utilize in practice. Only if the various moderators can be taken into account will such approaches be fully efficacious.

CONCLUSION

This research topic on whether lying is more effortful cognitively than truth-telling is meant to challenge long-held assumptions. Challenging assumptions is clearly a worthwhile scientific endeavor, and this collection of essays will doubtless enlighten the issue while raising a number of salient considerations.

In the process of addressing this assumption, however, let us not erect false dichotomies, straw-man arguments, or extreme positions that produce more heat than light. For example, the assertions by McCornack et al. (2014) that the differences between truth and deception should all be attributed to memory and information processing is serious overstatement, just as their assertion that current models of deception impute too much cognitive work to deceptive message production is an overly broad gloss. As with so many issues surrounding human cognition and behavior, simple answers are facile but inaccurate and will set our science back. The typology of 13 moderators I have proposed derives from modeling deception as a communication phenomenon, the properties of which can exacerbate or alleviate cognitive demands. The non-exhaustive collection of moderators includes: (1) sender memory demands, (2) sender motivation, (3) incentives and consequences, (4) discourse genre, (5) form of prevarication, (6) expected response length, (7) sanctioning of the deceit, (8) communication medium, (9) advance preparation, (10) recency of the incident/issue, (11) relationship among interlocutors (e.g., cooperative or adversarial), (12) relational familiarity, and (13) size of unit of analysis. I invite further formalization and empirical testing by other deception scholars to disentangle the effects of these significant moderators.

REFERENCES

Abe, N., and Greene, J. D. (2014). Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. *J. Neurosci.* 34, 10564–10572. doi: 10.1523/JNEUROSCI.0217-14.2014

Bereby-Meyer, Y., and Shalvi, S. (2015). Deliberate honesty. *Curr. Opin. Psychol.* 6, 195–198. doi: 10.1016/j.copsyc.2015.09.004

Bond, C. F., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234. doi: 10.1207/s15327957pspr1003_2

Braun, M., Van Swol, L. M., and Vang, L. (2015). His lips are moving: pinocchio effect and other lexical indicators of political deceptions. *Discourse Process.* 52, 1–20. doi: 10.1080/0163853X.2014.942833

Buller, D. B., and Aune, R. K. (1987). Nonverbal cues to deception among intimates, friends, and strangers. *J. Nonverb. Behav.* 11, 269–290. doi: 10.1007/BF00987257

Buller, D. B., and Burgoon, J. K. (1996a). Another look at information management: a rejoinder to McCornack, Levine, Morrison, and Lapinski. *Commun. Monogr.* 63, 92–98. doi: 10.1080/03637759609376377

Buller, D. B., and Burgoon, J. K. (1996b). Interpersonal deception theory. *Commun. Theory* 6, 203–242. doi: 10.1111/j.1468-2885.1996.tb00127.x

Buller, D. B., Burgoon, J. K., White, C., and Ebesu, A. S. (1994). Interpersonal deception: VII. behavioral profiles of falsification, concealment, and equivocation. *J. Lang. Soc. Psychol.* 13, 366–395. doi: 10.1177/0261927X94134002

Burgoon, J. K. (2015). Rejoinder to Levine, Clare et al.'s comparison of the Park–Levine probability model versus interpersonal deception theory: application to deception detection. *Hum. Commun. Res.* 41, 327–349. doi: 10.1111/hcre.12065

Burgoon, J. K., Blair, J. P., and Strom, R. (2008). Cognitive biases, modalities and deception detection. *Hum. Commun. Res.* 34, 572–599.

Burgoon, J. K., and Buller, D. B. (1996). Reflections on the nature of theory building and the theoretical status of interpersonal deception theory. *Commun. Theory* 6, 311–328. doi: 10.1111/j.1468-2885.1996.tb00132.x

Burgoon, J. K., and Buller, D. B. (2015). "Interpersonal deception theory: Purposive and interdependent behavior during deceptive interpersonal interactions," in

Engaging Theories in Interpersonal Communication, 2e, eds D. O. Braithwaite and P. Schrodt (Los Angeles, CA: Sage Publications), 349–362.

Burgoon, J. K., Buller, D. B., and Floyd, K. (2001). Does participation affect deception success? A test of the inter-activity effect. *Hum. Commun. Res.* 27, 503–534.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W., and Feldman, C. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Commun. Monogr.* 63, 50–69. doi: 10.1080/03637759609376374

Burgoon, J. K., and Floyd, K. (2000). Testing for the motivation impairment effect during deceptive and truthful interaction. *Western J. Commun.* 64, 243–267. doi: 10.1080/10570310009374675

Burgoon, J. K., and Langer, E. (1995). "Language, fallacies, and mindlessness-mindfulness," in *Communication Yearbook 18*, ed. B. Burleson (Newbury Park, CA: Sage Publication), 105–132.

Burgoon, J. K., Marett, K., and Blair, J. P. (2004). "Detecting deception in computer-mediated communication," in *Computers in Society: Privacy, Ethics and the Internet*, ed. J. F. George (Upper Saddle River, NJ: Prentice-Hall), 154–166.

Burgoon, J. K., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., et al. (2015). Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *J. Lang. Soc. Psychol.* doi: 10.1177/0261927X15586792

Burgoon, J. K., Nunamaker, J. F. Jr., and Metaxas, D. (2010). *Noninvasive Measurement of Multimodal Indicators of Deception and Credibility*. Final Report to the Defense Academy for Credibility Assessment. Tucson: University of Arizona.

Burgoon, J. K., Proudfoot, J. G., Wilson, D., and Schuetzler, R. (2014). Patterns of nonverbal behavior associated with truth and deception: illustrations from three experiments. *J. Nonverb. Behav.* 38, 325–354. doi: 10.1007/s10919-014-0181-5

Burgoon, J. K., Qin, T., Hamel, L., and Proudfoot, J. (2012). "Predicting veracity from linguistic indicators," in *Paper Presented to the Workshop on Innovation in Border Control (WIBC) at the European Intelligence and Security Informatics Conference (EISIC)*, Odense.

Christ, E. C., van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566. doi: 10.1093/cercor/bhn189

DePaulo, B. M., Ansfield, M. E., and Bell, K. L. (1996). lnterpersonal deception theory. *Commun. Theory* 6, 297–310. doi: 10.1111/j.1468-2885.1996.tb00131.x

DePaulo, B., and Kirkendol, S. E. (1989). "The motivational impairment effect in the communication of deception," in *Credibility Assessment*, ed. J. Yuille (Deurne: Kluwer), 51–70.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74

Dunbar, N. E., Jensen, M. L., Bessabarova, E., Burgoon, J. K., Bernard, D. R., Robertson, K. J., et al. (2014). Empowered by persuasive deception: the effects of power and deception on interactional dominance, credibility, and decision-making. *Commun. Res.* 41, 852–876. doi: 10.1177/0093650212447099

Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Kelley, K. M., Harrison, K. J., Adame, B., et al. (2013). Effects of veracity, modality and sanctioning on credibility assessment during mediated and unmediated interviews. *Commun. Res.* 40, 1–26.

Ennis, E., Vrij, A., and Chance, C. (2008). Individual differences and lying in everyday life. *J. Soc. Pers. Relat.* 25, 105–118. doi: 10.1177/0265407507086808

Fisher, R. P., and Geiselman, R. E. (1992). *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Charles C Thomas.

Frank, M. G., and Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *J. Appl. Commun. Res.* 31, 58–75. doi: 10.1080/00909880305377

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836. doi: 10.1093/cercor/13.8.830

Goldman-Eisler, F. (1958). Speech analysis and mental processes. *Lang. Speech* 1, 59–75.

Gougler, M., Nelson, R., Handler, M., Krapohl, D., Shaw, P., and Bierman, L. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph* 40, 194–305.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hartwig, M., and Bond, C. F. Jr. (2014). Lie detection from multiple cues: a meta-analysis. *Appl. Cogn. Psychol.* 28, 661–676. doi: 10.1002/acp.3052

Hopper, R., and Bell, R. A. (1984). Broadening the deception construct. *Q. J. Speech* 70, 288-302.

Horvath, F., Blair, J. P., and Buckley, J. P. (2008). The behavioural analysis interview: clarifying the practice, theory and understanding of its use and effectiveness. *Int. J. Police Sci. Manag.* 10, 101–118. doi: 10.1350/ijps.2008.10.1.101

Ito, A., Abe, N., Fujii, T., Ueno, A., Koseki, Y., Hashimoto, R., et al. (2011). The role of the dorsolateral prefrontal cortex in deception when remembering neutral and emotional events. *Neurosci. Res.* 69, 121–128. doi: 10.1016/j.neures.2010.11.001

Kellermann, K. (1992). Communication: inherently strategic and primarily automatic. *Commun. Monogr.* 59, 288–300. doi: 10.1080/03637759209376270

Knapp, M. L. (2008). *Lying and Deception in Human Interaction*. Boston, MA: Allyn and Bacon.

Levine, T. R., Shaw, A., and Shulman, H. C. (2010). Increasing deception detection accuracy with strategic questioning. *Hum. Commun. Res.* 36, 216–231. doi: 10.1111/j.1468-2958.2010.01374.x

Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., et al. (2010). Dorsolateral prefrontal cortex specifically processes general – but not personal –knowledge deception: multiple brain networks for lying. *Behav. Brain Res.* 211, 164–168. doi: 10.1016/j.bbr.2010.03.024

McCornack, S. A. (1997). "The generation of deceptive messages: laying the groundwork for a viable theory of interpersonal deception," in *Message Production: Advances in Communication Theory*, ed. J. O. Greene (Mahwah, NJ: LEA), 91–126.

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., and Zhu, X. (2014). Information manipulation theroy 2 a propositional theroy of deceptive discourse production. J. Lang. and Soc. Psychol., 33, 348–377. doi: 10.1177/0261927x14534656

Miller, G. R., and Stiff, J. B. (1993). *Deceptive Communication*. Thousand Oaks, CA: Sage Publcation.

Mullin, D. S., King, G. W., Saripalle, S. K., Derakhshani, R. R., Lovelace, C. T., and Burgoon, J. K. (2014). Deception effects on standing center of pressure. *Hum. Mov. Sci.* 38, 106–115. doi: 10.1016/j.humov.2014.08.009

O'Hair, H. D., Cody, M. J., and McLaughlin, M. L. (1981). Prepared lies, spontaneous lies, Machiavellianism and nonverbal communication. *Hum. Commun. Res.* 7, 325–339. doi: 10.1111/j.1468-2958.1981.tb00579.x

Porter, S., and ten Brinke, L. (2010). The truth about lies: what works in detecting high-stakes deception? *Legal Criminol. Psychol.* 15, 57–75. doi: 10.1348/135532509X433151

Rockwell, P., Buller, D. B., and Burgoon, J. K. (1997). The voice of deceit: refining and expanding vocal cues to deception. *Commun. Res. Rep.* 14, 451–459. doi: 10.1080/08824099709388688

Searcy, W. A., and Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton, NJ: Princeton University Press.

Shalvi, S., and De Dreu, C. K. W. (2014). Oxytocin promotes group serving dishonesty. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5503–5507. doi: 10.1073/pnas.1400724111

Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychol. Sci.* 23, 1264–1270. doi: 10.1177/0956797612443835

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853. doi: 10.1097/00001756-200109170-00019

Spence, S. A., Kaylor-Hughes, C., Farrow, T. F. D., and Wilkinson, I. D. (2008). Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception. *Neuroimage* 40, 1411–1418. doi: 10.1016/j.neuroimage.2008.01.035

Sporer, S. L., and Schwandt, B. (2006). Paraverbal indicators of deception: a meta-analytic synthesis. *Appl. Cogn. Psychol.* 20, 421–446. doi: 10.1002/acp.1190

Sporer, S. L., and Schwandt, B. (2007). Moderators of nonverbal indicators of deception: a meta-analytic synthesis. *Psychol. Public Policy Law* 13, 1–34. doi: 10.1037/1076-8971.13.1.1

Tabatabaeian, M., Dale, R., and Duran, N. (2015). Self-serving dishonest decisions can show facilitated cognitive dynamics. *Cogn. Process.* 16, 291–300. doi: 10.1007/s10339-015-0660-6

ten Brinke, L., and Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law Hum. Behav.* 36, 469–477. doi: 10.1037/h0093929

Turner, R. E., Edgley, C., and Olmstead, G. (1975). Information control in conversations: honesty is not always the best policy. *Kansas J. Speech* 11, 69–89.

Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practices*. West Sussex: John Wiley and Sons.

Vrij, A., Fisher, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci. (Regul. Ed.)* 10, 141–142. doi: 10.1016/j.tics.2006.02.003

Vrij, A., and Granhag, P. A. (2012). Eliciting cues to deception and truth: what matters are the questions asked. *J. Appl. Res. Mem. Cogn.* 1, 110–117. doi: 10.1016/j.jarmac.2012.02.004

Vrij, A., Semin, G. R., and Bull, R. (1996). Insight into behavior displayed during deception. *Hum. Commun. Res.* 22, 544–562. doi: 10.1111/j.1468-2958.1996.tb00378.x

Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Adv. Exp. Soc. Psychol.* 14, 1–59. doi: 10.1016/S0065-2601(08)60369-X

Zuckerman, M., and Driver, R. (1985). "Telling lies: verbal and nonverbal correlates of deception," in *Nonverbal Communication: An Integrated Perspective*, eds A. W. Siegman and S. Feldstein (Hillsdale, NJ: Erlbaum), 129–147.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 February 2016
doi: 10.3389/fpsyg.2016.00113

One-by-One or All-at-Once? Self-Reporting Policies and Dishonesty

Rainer M. Rilke[1], Amos Schurr[2]*, Rachel Barkan[2]* and Shaul Shalvi[3, 4]**

[1] Department of Corporate Development and Business Ethics, University of Cologne, Cologne, Germany, [2] Department of Business Administration, Ben Gurion University of the Negev, Beer-Sheva, Israel, [3] CREED, Faculty of Economics and Business, University of Amsterdam, Amsterdam, Netherlands, [4] Department of Psychology, Ben Gurion University of the Negev, Beer-Sheva, Israel

Organizational monitoring relies frequently on self-reports (e.g., work hours, progress reports, travel expenses). A "one-by-one" policy requires employees to submit a series of reports (e.g., daily or itemized reports). An "all-at-once" policy requires an overall report (e.g., an annual or an overview report). Both policies use people's self-reports to determine their pay, and both allow people to inflate their reports to get higher incentives, that is, to cheat. Objectively, people can cheat to the same extent under both reporting policies. However, the two policies differ in that the segmented one-by-one policy signals closer monitoring than the all-at-once policy. We suggest here that lie aversion may have a paradoxical effect on closer monitoring and lead people to cheat more. Specifically, reporting a series of segmented units of performance (allowing small lies) should lead to more cheating than a one-shot report of overall performance (that require one larger lie). Two surveys indicated that while people perceive the all-at-once policy as more trusting, they still expected people would be equally likely to cheat in both policies. An experiment tested the effects of the two reporting policies on cheating. The findings showed that contrary to the participants' intuition, but in line with research on lie aversion, the one-by-one policy resulted in more cheating than the all-at-once policy. Implications for future research and organization policy are discussed.

Keywords: dishonesty, behavioral ethics, monitoring, trust, lie aversion, justifications, organizational policy

OPEN ACCESS

Edited by:
Guy Hochman,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Celia Moore,
London Business School, UK
Lisa Dianne Ordonez,
University of Arizona, USA

****Correspondence:***
Rainer M. Rilke
rainer.rilke@uni-koeln.de;
Amos Schurr
samos@bgu.ac.il;
samos@som.bgu.ac.il;
Rachel Barkan
barkanr@som.bgu.ac.il;
Shaul Shalvi
s.shalvi@uva.nl

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *14 May 2015*
Accepted: *20 January 2016*
Published: *17 February 2016*

Citation:
Rilke RM, Schurr A, Barkan R and Shalvi S (2016) One-by-One or All-at-Once? Self-Reporting Policies and Dishonesty. Front. Psychol. 7:113. doi: 10.3389/fpsyg.2016.00113

INTRODUCTION

Honesty and trust are cornerstones of organizational success. For instance, Watson Wyatt's Work USA 2002 survey indicated the 3-year total return to shareholders was almost three times higher in companies characterized by high levels of honesty and trust than in companies characterized by low levels of honesty and trust. Decades of organizational research back up this example, teaching us that honesty and trust are important to both employers and employees (McGregor, 1960; Jones, 1991; Murphy, 1993; Moore and Gino, 2015). Honesty and trust are associated with higher levels of cooperation, better performance, proactive actions, effective management, and organizational growth (e.g., Jones and George, 1998; De Cremer et al., 2001; Dirks and Ferrin, 2001; Tyler, 2003; Cook et al., 2005). In addition, research suggests that ethical behavior elicits intrinsic incentives such as satisfaction and a sense of self-dignity (Peer et al., 2014; Moore and Gino, 2015).

Clearly, maintaining high ethical standards and fostering trust is advantageous for organizations. However, the combination of honesty and trust is easier preached than achieved. In fact, research suggests a counter-productive tradeoff exists between the two constructs. Specifically, close monitoring and increased enforcement are effective means to increase honesty, but are considered detrimental to trust (Kirchler, 2007; Kirchler et al., 2008). Lowering the levels of monitoring may boost trust, but—as rational economic analysis and behavioral research show—such leniency frequently leads to a gradual deterioration of ethical standards and an increase in dishonest behavior (Becker, 1974; Kirchler, 2007; Kirchler et al., 2008; Gino and Bazerman, 2009). Thus, designing an organizational environment that encourages both honest behavior and trust is a challenge, and choosing optimal policies is not straightforward.

A common organizational solution lets employees monitor and report their own performance. Taking self-reports at face value, organizations signal that they trust their employees and expect honest reports of true performance in return. An optimistic view suggests that trusting policies can foster loyalty, increase productivity, boost satisfaction, and reduce turnover (e.g., Shockley-Zalabak et al., 2000; Parker et al., 2006). A more pessimistic take is that self-report policies may tempt employees to inflate their performance levels and cheat at the expense of the organization.

Specific reporting procedures and especially the resolution level of self-reports, reflect these different takes on self-reported performance. Some organizations require regular segmented reports (e.g., hours worked, number of items produced, quality control measures; here dubbed the "one-by-one policy"). Other organizations require overall reports (e.g., summarizing a work project, expense reimbursements; here dubbed the "all-at-once" policy). By its nature, the micro-management approach of the one-by-one policy provides a higher level of monitoring compared to the relatively more macro-management approach of the all-at-once policy. In the current work, we compare the two reporting policies, and examine whether the difference between the policies is merely a semantic nuance, or whether it leads to different responses and affects the level of honesty.

Objectively, people can exploit both policies to the same extent. For example, consider a company representative reporting the number of customers that were interested in a new service he offered. Suppose that on 5 consecutive days, the numbers of interested customers were 1, 2, 3, 4, and 5. The representative can inflate daily reports of performance (e.g., reporting 2, 3, 4, 5, and 6) or inflate an overall report (e.g., report 20 instead of 15). Thus, from a rational economic analysis the two policies should lead to similar levels of cheating. However, Research on lie aversion shows that people justify small lies more easily than big lies. To avoid the adversity of being "real" liars, people tend to restrict their own dishonesty to a level they can justify (e.g., Ayal et al., 2015; Shalvi et al., 2015). Thus, rather than cheating to the maximal extent (for maximal profit), people tend to cheat "only by a little" to benefit from cheating but still maintain a sense of morality (Mazar et al., 2008; Shalvi et al., 2011a). Empirical evidence indicates that even when participants are guaranteed they will not be caught, punished, or even identified as cheaters, they still exhibit lie aversion (Gneezy, 2005; Lundquist et al., 2009; Shalvi et al., 2011b; Hilbig and Hessler, 2013; Weisel and Shalvi, 2015). Applying the reasoning of lie aversion to the abovementioned hypothetical example suggests the company's representative will feel more comfortable telling five small lies than one big lie. Accordingly, people will cheat more when they submit a series of small reports in the one-by-one policy, and will restrain themselves when they submit a single overall report in the all-at-once policy. This is the possibility we test here.

Interestingly, there are two ways in which the mean level of cheating can be higher in the one-by-one compared to the all-in-once setting. One option is that more people may be tempted to lie just a bit in the one-by-one setting compared to the all-at-once setting. If this is true, we should observe that the distribution of reported performance in the one-by-one distribution is shifted to the right indicating more people reported higher outcomes. A second option is that a similar proportion of people will lie in both settings, but lies will be larger in the one-by-one policy. The latter option bears a resemblance to the "what-the hell" effect—that is, once people cave to lying, they lie by a lot (Mazar et al., 2008; Mead et al., 2009; Ariely, 2012). We test these two possibilities.

A recent paper provides initial support for our argument that a one-by-one setting should lead to more lying than an all-at-once setting. In this work, Schurr et al. (2012) examined the effect of different choice procedures on dishonesty. In one of their experiments participants played a 20-questions trivia game. However, instead of answering the questions, participants were first shown the correct answer and then asked to report whether that was the answer they had in mind. Participants earned money every time they reported they had the correct answer in mind. Obviously, participants could lie to earn more money, and the experimenters had no way to detect lies or liars. Importantly, participants earned more money for difficult questions [e.g., Samuel Langhorne Clemens is better known as: (a) Rudyard Kipling; (b) Edgar Allan Poe; (c) *Mark Twain*; (d) Oscar Wilde] and less money for easy questions [e.g., "The Portrait of Dorian Gray" is a novel by: (a) Rudyard Kipling; (b) Edgar Allan Poe; (c) Mark Twain; or (d) *Oscar Wilde*]. In one condition, participants were asked to choose between an easy and a difficult question before each trial. In another condition, participants decided ahead of time on the number of easy and difficult questions they wanted to solve in 20 trials. When facing a sequence of 20 temptations to cheat (as compared to a single temptation), participants caved in and chose more difficult (i.e., profitable) questions, to which they frequently reported they had the correct answer in mind. Note, however, that in both conditions, the task was identical (answering a sequence of trivia questions). Thus, rather than comparing the effect of reporting policies, this study primarily examined the effect of one planned choice vs. repeated ongoing choices on ethicality.

We now turn to report two surveys and an experiment that aimed to answer the question: Which reporting policy is more effective in encouraging honest self-reports—the one-by-one or the all-at-once?

The surveys tested people's intuitions regarding the two reporting policies. Participants of the first survey thought that the macro-management all-at-once reporting policy conveys more trust from the organization than the micro-management one-by-one policy. Independent of that trust, other participants completing the second survey expected the levels of dishonesty employees will engage in should be the same for the two reporting policies. That is, whereas people perceived the all-at-once policy to demonstrate higher level of trust in employees from the organization's perspective, this trust does not translates to people's prediction regarding employees' likelihood to behave unethically. Next, we report an experiment providing a direct comparison between the one-by-one and all-at-once reporting policies. The experimental task simulated employees self-reporting their performance to their employer for payment. The experimental findings indicate the one-by-one policy led to more cheating than the all-at-once policy. Implications for future research and for organization policy are discussed.

PEOPLE'S INTUITIONS

As a first step, we assessed the extent to which people have a clear and consensual intuitions regarding the one-by-one and all-at-once reporting policies. This is important as policies are designed based on what people believe the state of the world is. In the first survey, we asked a group of participants which of the two policies conveys more trust in employees' honesty. In a second survey we employed we asked a different group of participants whether employees are more likely to cheat in one of the two policies.

Intuition Regarding Trust

Materials and Method

For this survey, we recruited 93 participants to complete a paid online questionnaire via the Amazon Mechanical Turk website (46 females, $M_{age} = 36.18$, $SD_{age} = 10.41$).

The questionnaire described two fictional companies: Company A utilizes a one-by-one reporting policy and requires regular itemized reports of performance. Company B utilizes an all-at-once reporting policy requiring an integrated overall report of performance. Referring to five common types of performance (working hours, travel expenses, overtime, work progress, and calling in sick), we asked participants to state which of the two companies is more likely to trust their employees to report honestly. For example, the question regarding travel expenses read: "Company A requires their employees to report each of their travel expenses (food, hotels, taxis etc.) on separate forms. Company B requires their employees to report all their travel expenses (food, hotels, taxis, etc.) on just one form." For each question, participants chose one of three responses that read: "[1] Employees in Company A are more trusted by the organization to report honestly; [2] Employees in Company B are more trusted by the organization to report honestly; [3] Employees in Companies A and B are trusted to report honestly to the same extent by their organizations."

Results and Discussion

The relative frequencies distributions for each of the five questions showed that most participants felt that the all-at-once reporting policy (of Company B) conveys more trust than the one-by-one reporting policy (of Company A). A pooled distribution summarizes the general intuition (see **Figure 1**). A small proportion of the participants (13%) thought that the one-by-one policy of company A reflect more trust in the employees, a small proportion of participants (15%) thought both policies reflect the same level of trust, and the vast majority of the participants (72%) stated that the all-at-once reporting policy reflects more trust in the employees. We used effect coding (-1 = trust is more likely in Company A; 0 = trust is equally likely in both companies; 1 = trust is more likely in Company B). In line with the frequency distribution, the average of the pooled responses was significantly greater than zero ($M = 0.59$, $SD = 0.71$) $t_{(92)} = 8.02$ $p < 0.0001$.

Intuition Regarding Cheating

Materials and Methods

For a second survey, we recruited an additional 102 participants to complete a paid online questionnaire via the Amazon Mechanical Turk website (51 females, $M_{age} = 35.17$, $SD_{age} = 11.91$).

The survey was identical to the one described above, but this time we asked participants to state in which of the two companies employees are more likely to inflate self-reports. Participants chose one of three responses that read: "[1] Employees in Company A are more likely to inflate their reports; [2] Employees in Company B are more likely to inflate their reports; [3] Employees in Companies A and B are as likely to inflate their reports."

Results and Discussion

The relative frequencies distributions for each of the five questions showed that most participants expected inflated self-reports to be equally likely whether the reporting policy was one-by-one (i.e., Company A) or all-at-once (i.e., Company B). A pooled distribution summarizes the general intuition (see **Figure 2**). A small proportion of the participants (13%) predicted

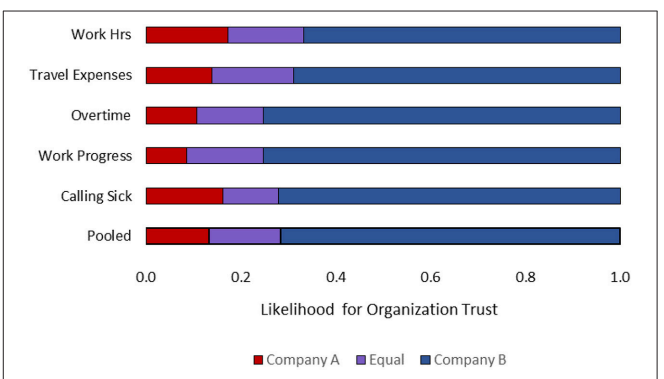

FIGURE 1 | Relative frequencies of participants' ascribed organizational trust in employees per company.

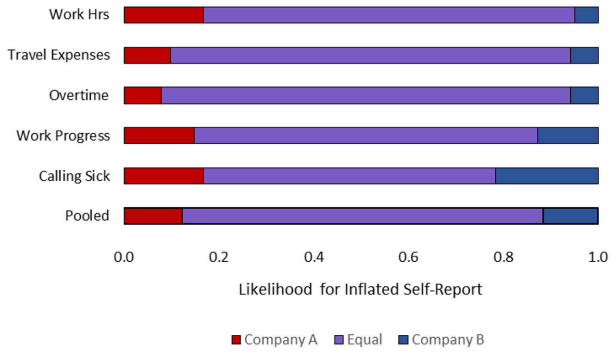

FIGURE 2 | Relative frequencies of participants' intuitions regarding the effect of reporting policy on inflated self-reports in the workplace.

that the one-by-one policy of company A would lead to more cheating, a small proportion of participants (10%) thought the all-at-once policy would lead to more cheating, and the vast majority of the participants (77%) stated that the reporting policy would not influence the rate of inflated self-reports. We used effect coding (-1 = cheating is more likely in Company A; 0 = cheating is equally likely in both companies; 1 = cheating is more likely in Company B). In line with the frequency distribution, the average of the pooled responses was practically zero ($M = 0.03$, $SD = 0.47$) $t_{(101)} = 0.62$ $p = 0.534$.

Thus, whereas participants perceived the macro-management all-at-once policy as more trusting than the micro-management one-by-one policy, this intuition did not translate to assuming that people will lie more in the one-by-one policy. We next report an experiment that compared the effect of the two reporting policies on actual cheating behavior in a controlled laboratory experiment.

EXPERIMENT

The experiment reported below provides a direct comparison of the one-by-one and all-at-once reporting policies. The experiment was designed to test if the different procedures of the one-by-one and all-at-once reporting policies affect the level of dishonesty.

Eliciting Cheating with a Trivia Game

To simulate a work setting and allow participants to earn money solely on the basis of self-reports, we adapted the trivia game paradigm (Schurr et al., 2012). The game included two rounds of 20 questions each[1]. The first round was entitled "practice" and did not involve incentives. The second round was entitled "test" and involved incentives. In the "test" round, participants earned a fixed payment each time they stated they had the correct answer in mind. Note that performance level could not improve between the two rounds (i.e., trivia games are based on existing knowledge and each question was presented only once). Thus, the practice round served to establish baseline performance,[2] and any improvement in the test round represented participants' inflating performance to earn more money. To simulate the two reporting policies, in the one-by-one condition, we presented participants with a series of 20 separate trials. In each trial, participants answered a single trivia question. In the all-at-once condition, we presented participants with a list of 20 trivia questions in a single trial. Participants solved all the 20 trivia questions in this one trial and submitted a 20-line report of their performance (see Appendix 1 in Supplementary Material).

Materials and Methods

We recruited 96 participants (43 females, $M_{age} = 23.53$, $SD_{age} = 4.89$) via an online web system to participate in an experiment at the Cologne Laboratory of Economic Research. We compensated participants with a fixed show-up fee of €5 and an added bonus contingent on their earnings in the test round of the trivia task (€0–€4). The experiment lasted about 30 min.

Upon arrival, participants were seated in computer cubicles that ensured privacy and were randomly allocated to one of the two reporting-policy conditions. In the one-by-one condition, in each trial, a single question was displayed on screen. Participants were asked to think about the answer, keep it in mind, and click a button to reveal the correct one. Participants then clicked a Yes/No button to report whether the answer they had in mind was correct. Participants repeated the same procedure for each of the 20 questions (see Appendix 1 in Supplementary Material). In the all-at-once condition, a list of all 20 questions appeared on the screen. Participants worked through the list (in any order they saw fit). For each question, participants were asked to think about the answer, click a button to reveal the correct one, and then click the Yes/No button to report whether they had the correct answer in mind. Participants could edit and change their responses before they submitted their overall reports (see Appendix 1 in Supplementary Material). In each condition, participants first completed a practice round (without incentives) and then completed a test round (with incentives). In the test round, participants earned €0.2 each time they reported they had the correct answer in mind.

Results and Discussion

The critical measure was the difference in reported performance between the test and practice rounds. As explained above, any improvement from the baseline performance represented cheating for monetary profit.

The findings showed that in both conditions, participants tended to inflate their performance in the test round to earn money. Importantly, participants were much more likely to inflate their performance reports in the one-by-one condition

[1]We selected 40 trivia questions for the practice and test rounds from a pool of 200 questions that were pretested. The level of difficulty was controlled and was comparable within and between rounds.

[2]One reason to inflate performance has to do with impression management and social desirability, in that no one wants to admit ignorance. Thus, the baseline performance that participants report in the practice round reflects a composite of their true performance and self-enhancement. The incentives in the test round present an additional temptation to cheat—this time for monetary profit. Thus, any "improvement" in the second round isolates the component of dishonesty that is harnessed to increase monetary profit.

(improving by 10.74%) than in the all-at-once condition (improving by 3.6%, see **Table 1**).

We submitted the overall performance reports in the practice and test rounds to a repeated measures ANOVA with reporting policy as a between-subject variable. The effect of Round was significant $F_{(1, 94)} = 16.96$, $p < 0.0001$ partial $\eta^2 = 0.15$. The main effect of Policy was not significant $F_{(1, 94)} < 0$, *ns*. The interaction between Round and Policy was marginally significant $F_{(1, 94)} = 3.91$, $p = 0.051$, partial $\eta^2 = 0.04$ indicating that in general participants tended to inflate performance more in the one-by-one condition than in the all-at-once condition. Simple effects analysis revealed significant improvement (compared to the null hypothesis that assumes no improvement) in the one-by-one [$t_{(46)} = 3.95$, $p < 0.0001$] but not in the all-at-once setting [$t_{(48)} = 1.67$, $p = 0.105$].

We further tested the effect using non-parametric tests. The baseline performance in the practice round differed slightly between the two experimental conditions, but the difference was not significant ($Z = 1.02$, $p = 0.307$, Mann-Whitney *U*-test). Participants' self-reports indicated improved (i.e., inflated) performance in the test round. A Wilcoxon signed rank test showed that this improvement was significant in the one-by-one condition ($Z = 3.473$, $p = 0.0005$), but was not significant in the all-at-once condition ($Z = 1.51$, $p = 0.130$). A direct comparison between the two experimental conditions indicated that improvement (i.e., inflated performance) was almost three times larger in the one-by-one condition than in the all-at-once condition ($Z = 2.14$, $p = 0.032$, Mann-Whitney *U*-test).

Thus, in line with lie aversion, the findings showed that the one-by-one policy led to more cheating, whereas the all-at-once policy resulted in considerable self-restraint.

In a follow up analysis we examined behavior at the individual level. Whereas, our design does not allow to determine if a certain participant lied or not, the "improvement" that a participant reports allows us to assess the likelihood of dishonesty. Specifically, in the Trivia paradigm, if a participant reports the same level of performance in the practice and test rounds, there is high likelihood of honesty. Negative improvement scores (i.e., lower performance in the test round compared to practice) suggest that honesty was chosen over the possibility to earn money (and at the personal cost of admitting ignorance). In contrast, reporting better performance in the test round, suggests it is likely that performance was falsely inflated to earn more money.

TABLE 1 | Self-reported performance on the trivia game.

	N (% female)	Average correct answers reported		"Improvement" average difference
		Practice phase	Test phase	
One-by-one	47 (53%)	14.06	15.57	1.51* (*SD* = 2.62)
All-at-once	49 (41%)	14.65	15.18	0.53 (*SD* = 2.22)

**The mean difference was significantly different from zero at the p < 0.01 level on the two-tailed Wilcoxon signed rank test.*

Figure 3 shows the relative frequency distributions of "improvement" scores in the two experimental conditions. As can be seen, negative to zero improvement scores (i.e., high likelihood of honesty) were more frequent in the all-at-once reporting policy. In contrast, high improvement scores (i.e., high likelihood of dishonesty) were more than twice as likely in the one-by-one policy. The difference between the frequency distributions was marginally significant $\chi^2_{(4)} = 8.81$, $p = 0.066$. The pattern lends further support to the idea that one-by-one reporting policy is more likely to facilitate dishonesty than all-at-once reporting policy.

GENERAL DISCUSSION

Self-reporting policies aim to apply organizational monitoring and encourage trust at the same time. Here, we compared two specific reporting policies. One policy, entitled one-by-one, requires employees to report separate segments of their performance. Another policy, entitled all-at-once, allows employees to submit an integrated overall report. Objectively, employees could exploit both policies and inflate self-reports to the same extent. Indeed, two surveys revealed that while people perceive the all-at-once policy as more trusting, they still expected people would be equally likely to cheat in both policies. Our results demonstrate however that people lie more in a one-by-one procedure than in an all-at-once procedure.

An analysis at the individual level indicated that participants were more likely to resist temptation and be honest when they were asked to provide a single report of their performance. In contrast, participants were more likely to inflate performance to a large extent when they provided a sequence of segmented reports. The finding is in line with the idea that repeated reports in the one-by-one policy make it harder for people to resist the temptation to cheat, and that once they cave in to the temptation to lie, they lie by a lot.

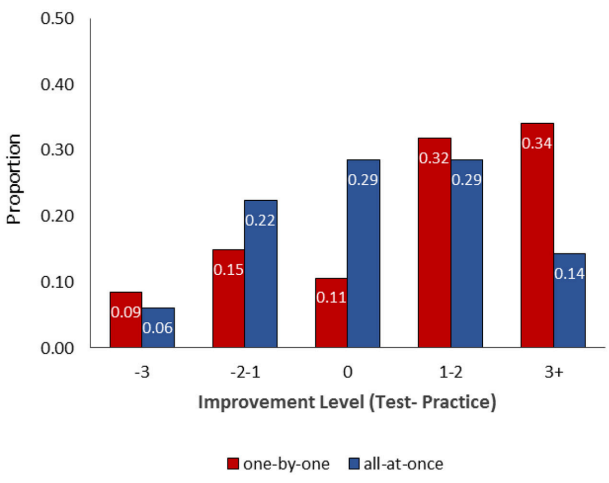

FIGURE 3 | Frequency distributions of participants' "improvement" scores in the one-by-one and all-at-once reporting policies.

As always, generalization of experimental findings should be cautious, and subject to the accumulation of research that offers replications of the effect as well as identifying boundary conditions. For example, Desai and Kouchaki (2015) recently reported a study that seemed to favor the one-by-one policy over the all-at-once policy. In this study, an experimenter contacted garage mechanics to obtain an estimate for changing the brake pads of a car. The findings showed that in this set-up, mechanics inflated costs less when they provided separate estimates for two different aspects of the job (i.e., parts, labor), than when they provided an overall cost estimate. The authors offered that specific estimates elicited higher accountability requiring mechanics to be able to justify their quotes (Desai and Kouchaki, 2015, study 7). It is difficult to establish a direct comparison between offering a price quote to a client and reporting performance to an employer. Still, the finding raises an interesting question regarding a potential difference between one-time task and repeated tasks. For example, what would have happened, if the mechanics had to provide price quotes repeatedly to more clients? Would the segmented quote still be lower than the overall quote for the 10th and 20th clients? This question sets an interesting direction for future research.

On a theoretical level, one-by-one and all-at-once reporting procedures can be used to test possible interactions between accountability (e.g., Tetlock, 1992; Lerner and Tetlock, 1999; Desai and Kouchaki, 2015) and lie aversion (e.g., Gneezy, 2005; Lundquist et al., 2009; Shalvi et al., 2011b; Hilbig and Hessler, 2013). The two constructs could probably be combined to work in the same direction and encourage honesty. In the setting we examined, however, we suspect that accountability and lie aversion may have operated against each other. To wit, the one-by-one itemized report may have elicited a sense of accountability and increased the need to justify one's actions. Note however, that in our experiment, such reporting procedure involved small lies that disappeared from the screen at the end of each trial. The combination of an increased need to justify one's action and small lies that are easily forgotten may have led to a paradoxical outcome that minimized the psychological cost of guilt and facilitated cheating. More research is needed to examine and fully establish the ways in which lie aversion and accountability interact, as well as the contextual factors that may determine their joint effect.

It is worth noting that the experimental task we employed provides only direct measures of self-reported improvement rather than explicit measures of cheating. While we assume that improvement in the trivia tasks points at high likelihood of cheating—it is of course possible that some participants indeed improved in their performance levels in the test round. We chose this experimental task because it has high external validity (as people often do not know if a person is cheating or not, just receive indirect indications for such possibility). Future research may benefit from experimental tasks that allow to trace individual's dishonesty more directly. Such future work can further explore if individual differences in relevant parameters such as gender, moral disengagement, or moral identity, may moderate any of the observed effects. This would increase our understanding of the patterns identified here.

On an applied note, the findings are also important for the development of organizational monitoring policies that aim to prompt both honesty and trust. Indeed, in a survey we found that people consider organizations implementing an all-at-once policy to be more trusting of their employees' honesty than organizations implementing the one-by-one policy. Our experiment suggests that close monitoring policy might lead people to discount segmented transgressions as minor and negligible and thus result in lower honesty levels. Our findings offer an optimistic view showing that the more trusting policy led to more honest behavior.

Many organizational activities can be categorized as one-by-one or all-at-once tasks. For example, many people sort out their incoming emails by automatically placing them in multiple folders, which in turn get packed with many items waiting to be dealt with. Does the number of folders affect people's responses to those emails? People also use reporting systems to report different purchase orders, submit expenses reports, tax reports, and so forth. Our findings suggest that the effect of reporting procedure on honesty is not trivial and policy makers should consider carefully the reporting procedures and even the display format of reporting systems to design settings that encourage ethical conduct.

ETHIC STATEMENT

All studies were approved by the Institutional Review Boards at the university of Cologne. All participants read and signed an informed consent before the studies.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to Theresa Eyerund for her help in conducting the experiments. Financial support was received from the Deutsche Forschungsgemeinschaft through grant "TP3 Design of Incentives Schemes within Firms: Bonus Systems and Performance Evaluations" (sub-project of the DFG-Forschergruppe "Design and Behavior"), from a Leibniz-Award to Axel Ockenfels, and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement ERC-StG-637915 to SS), all are gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00113

REFERENCES

Ariely, D. (2012, May 26). Why we lie. *WSJ. com*, pp. 531–548.

Ayal, S., Gino, F., Barkan, R., and Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspect. Psychol. Sci.* 10, 738–741. doi: 10.1177/1745691615598512

Becker, G. S. (1974). "Crime and punishment: an economic approach," in *Essays in the Economics of Crime and Punishment* (Cambridge, MA: NBER), 1–54.

Cook, K. S., Hardin, R., and Levi, M. (2005). *Cooperation Without Trust?* New York, NY: Russell Sage Foundation.

De Cremer, D., Snyder, M., and Dewitte, S. (2001). "The less I trust, the less I contribute (or not)?" The effects of trust, accountability and self−monitoring in social dilemmas. *Eur. J. Soc. Psychol.* 31, 93–107. doi: 10.1002/ejsp.34

Desai, S. D., and Kouchaki, M. (2015). Work-report formats and overbilling: how unit-reporting vs. cost-reporting increases accountability and decreases overbilling. *Organ. Behav. Hum. Decis. Process.* 130, 79–88. doi: 10.1016/j.obhdp.2015.06.007

Dirks, K. T., and Ferrin, D. L. (2001). The role of trust in organizational settings. *Organ. Sci.* 12, 450–467. doi: 10.1287/orsc.12.4.450.10640

Gino, F., and Bazerman, M. H. (2009). When misconduct goes unnoticed: the acceptability of gradual erosion in others' unethical behavior. *J. Exp. Soc. Psychol.* 45, 708–719. doi: 10.1016/j.jesp.2009.03.013

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Hilbig, B. E., and Hessler, C. M. (2013). What lies beneath: how the distance between truth and lie drives dishonesty. *J. Exp. Soc. Psychol.* 49, 263–266. doi: 10.1016/j.jesp.2012.11.010

Jones, G. R., and George, J. M. (1998). The experience and evolution of trust: implications for cooperation and teamwork. *Acad. Manage. Rev.* 23, 531–546.

Jones, T. M. (1991). Ethical decision making by individuals in organizations: an issue-contingent model. *Acad. Manage. Rev.* 16, 366–395.

Kirchler, E. (2007). *The Economic Psychology of Tax Behaviour*. Cambridge, MA: Cambridge University Press.

Kirchler, E., Hoelzl, E., and Wahl, I. (2008). Enforced versus voluntary tax compliance: the "slippery slope" framework. *J. Econ. Psychol.* 29, 210–225. doi: 10.1016/j.joep.2007.05.004

Lerner, J. S., and Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychol. Bull.* 125, 255. doi: 10.1037/0033-2909.125.2.255

Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. *J. Econ. Behav. Organ.* 70, 81–92. doi: 10.1016/j.jebo.2009.02.010

Mazar, N., Amir, O., and Ariely, D. (2008). More ways to cheat-expanding the scope of dishonesty. *J. Mark. Res.* 45, 651–653.

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., and Ariely, D. (2009). Too tired to tell the truth: self-control resource depletion and dishonesty. *J. Exp. Soc. Psychol.* 45, 594–597. doi: 10.1016/j.jesp.2009.02.004

McGregor, D. (1960). *The Human Side of Enterprise*. New York, NY: McGraw-Hill.

Moore, C., and Gino, F. (2015). Approach, ability, aftermath: a psychological process framework of unethical behavior at work. *Acad. Manag. Ann.* 9, 235–289. doi: 10.1080/19416520.2015.1011522

Murphy, K. R. (1993). *Honesty in the Workplace.* Belmont, CA: Thomson Brooks/Cole Publishing Co.

Parker, S. K., Williams, H. M., and Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *J. Appl. Psychol.* 91, 636. doi: 10.1037/0021-9010.91.3.636

Peer, E., Acquisti, A., and Shalvi, S. (2014). "I cheated, but only a little": partial confessions to unethical behavior. *J. Pers. Soc. Psychol.* 106, 202–217. doi: 10.1037/a0035392

Schurr, A., Ritov, I., Kareev, Y., and Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgm. Decis. Mak.* 7, 679–688. Available online at: http://journal.sjdm.org/12/12916/jdm12916.html

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Shalvi, S., Handgraaf, M. J., and De Dreu, C. K. (2011a). People avoid situations that enable them to deceive others. *J. Exp. Soc. Psychol.* 47, 1096–1106. doi: 10.1016/j.jesp.2011.04.015

Shalvi, S., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011b). Ethical manoeuvring: why people avoid both major and minor lies. *Br. J. Manage.* 22, S16–S27. doi: 10.1111/j.1467-8551.2010.00709.x

Shockley-Zalabak, P., Ellis, K., and Winograd, G. (2000). Organizational trust: what it means, why it matters. *Organ. Dev. J.* 18, 35–48. Available online at: http://search.proquest.com/docview/197985640?accountid=14546

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: toward a social contingency model. *Adv. Exp. Soc. Psychol.* 25, 331–376. doi: 10.1016/S0065-2601(08)60287-7

Tyler, T. R. (2003). Trust within organisations. *Pers. Rev.* 32, 556–568. doi: 10.1108/00483480310488333

Weisel, O., and Shalvi, S. (2015). The collaborative roots of corruption. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10651–10656. doi: 10.1073/pnas.1423035112

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 02 June 2016
doi: 10.3389/fpsyg.2016.00734

When Lying Feels the Right Thing to Do

Sophie Van Der Zee[1,2,3], Ross Anderson[1] and Ronald Poppe[4]*

[1] Computer Laboratory, University of Cambridge, Cambridge, UK, [2] Department of Medical and Sport Sciences, University of Cumbria, Lancaster, UK, [3] Networked Organisations, TNO, Rijswijk, Netherlands, [4] Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

Fraud is a pervasive and challenging problem that costs society large amounts of money. By no means all fraud is committed by 'professional criminals': much is done by ordinary people who indulge in small-scale opportunistic deception. In this paper, we set out to investigate when people behave dishonestly, for example by committing fraud, in an online context. We conducted three studies to investigate how the rejection of one's efforts, operationalized in different ways, affected the amount of cheating and information falsification. Study 1 demonstrated that people behave more dishonestly when rejected. Studies 2 and 3 were conducted in order to disentangle the confounding factors of the nature of the rejection and the financial rewards that are usually associated with dishonest behavior. It was demonstrated that rejection in general, rather than the nature of a rejection, caused people to behave more dishonestly. When a rejection was based on subjective grounds, dishonest behavior increased with approximately 10%, but this difference was not statistically significant. We subsequently measured whether dishonesty was driven by the financial loss associated with rejection, or emotional factors such as a desire for revenge. We found that rejected participants were just as dishonest when their cheating did not led to financial gain. However, they felt stronger emotions when there was no money involved. This seems to suggest that upon rejection, emotional involvement, especially a reduction in happiness, drives dishonest behavior more strongly than a rational cost-benefit analysis. These results indicate that rejection causes people to behave more dishonestly, specifically in online settings. Firms wishing to deter customers and employees from committing fraud may therefore benefit from transparency and clear policy guidelines, discouraging people to submit claims that are likely to be rejected.

Keywords: deception, dishonesty, rejection, insurance fraud, MTurk

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Jennifer Jordan,
University of Groningen, Netherlands
Tehila Kogut,
Ben-Gurion University of the Negev, Israel

****Correspondence:***
Sophie Van Der Zee
sophie.vanderzee@tno.nl

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *16 September 2015*
Accepted: *03 May 2016*
Published: *02 June 2016*

Citation:
Van Der Zee S, Anderson R and Poppe R (2016) When Lying Feels the Right Thing to Do. Front. Psychol. 7:734. doi: 10.3389/fpsyg.2016.00734

INTRODUCTION

Fraud is a pervasive and expensive problem: estimates of the cost of fraud vary from under 1% of GDP to over 10%, with the largest recent estimate of global fraud costs put at £7.22 trillion, or one seventh of global GDP (Gee and Button, 2013). A more modest estimate suggests that fraud is costing UK citizens approximately £1100 a year each (Centre for Counter Fraud Studies, 2014). According to the UK Fraud Act, fraud consists of "dishonestly making a false representation with the intent to make a gain for oneself or another, or to cause loss to someone else": in short, "an act

of deception intended for personal gain or to cause a loss to another party," as the Serious Fraud Office summarizes the law.

In this paper, we focus on deception in online settings. A detailed study suggests that, while 'technical' offenses such as payment card fraud, online banking fraud, and Internet fraud have an annual cost of several 10s of pounds per citizen per year, online versions of traditional offenses such as tax and welfare fraud costs each citizen of a developed country 100s of pounds a year (Anderson et al., 2012). And these are just the financial costs. When the fraud victims are persons rather than institutions, they can also experience negative psychological consequences, increased physical and mental health issues and damage to their relationships (Button et al., 2014). Motivated by the high direct and indirect costs of fraud, a wide range of countermeasures have been introduced (Alanezi and Brooks, 2014; Centre for Counter Fraud Studies, 2014). These include increasing security and surveillance, increasing awareness amongst potential victims and calling for more vigorous prosecution of fraudsters (Anderson et al., 2012; Purkait, 2012); they unfortunately also include measures such as blaming fraud victims for their misfortune (Cross, 2013). Although these measures may be rational from the viewpoint of the actors who introduce them, they do not always take the more irrational side of human nature into account.

Not all fraud is committed by 'professional criminals,' that is by individuals who earn their living through committing offenses. Instead, both real-world case studies and experimental research have shown that very few people lie and cheat to a pathological extent (for an overview of experiments, see Ariely, 2012); instead the majority of people are 'opportunistic fraudsters' who lie and cheat a little. Padded expenses, inflated insurance claims, refunds for goods wrongly said to have been defective, overtime payments for tea breaks; the world of trade, commerce and employment are beset with dishonest behavior. In a series of experiments involving participants who tried to solve as many matrix puzzles as they could within several minutes (Mazar et al., 2008; Gino et al., 2009), it was consistently found that people overstate their achievements by about 60% if they have the chance.

To prevent people from deceiving, we need to understand the factors that cause people to behave dishonestly. The deterrence of deception lies at the heart of most fraud countermeasures. In the last decade, vibrant research on the deterrence of deceit and dishonesty has emerged. This paper is aimed at a better understanding of when dishonest behavior occurs. Previous research has demonstrated that the extent to which people behave dishonestly is affected by several factors including individual differences, context and the environment. Creativity is one example of an individual difference that is linked to dishonesty. Creative people are more likely to behave dishonestly (Gino and Ariely, 2012), and are also better at it (Vrij, 2008). While individual differences can cause some people to be more dishonest than others, situational factors can increase the likelihood still further. Dishonesty tends to fluctuate during the day (Kouchaki and Smith, 2014). Throughout the day, people become more depleted, lowering their moral awareness and self-control. Therefore, dishonesty tends to increase as the day progresses. People tend to be more dishonest under certain circumstances, for example when pursuing a goal (Schweitzer et al., 2004), or in the presence of a bad example such as counterfeit goods (Gino et al., 2010). The behavior of other people can also influence a person's tendency to act dishonestly. For example, people are more likely to behave dishonestly when witnessing an in-group member behaving dishonestly (Gino et al., 2009), and when feeling socially rejected (Kouchaki and Wareham, 2015). The latter effect was mediated by physiological arousal, a finding that is in line with previous research suggesting that feelings of anxiety can promote dishonesty (Kouchaki and Desai, 2014). Therefore, non-social situations that elicit feelings of anxiety may also elicit dishonest behavior.

These situational factors have in common that they can be used to justify unethical behavior. Blasi (1980) identified a psychological gap that can emerge when people's moral understanding and their moral actions are not aligned. A possible explanation for the mental processes that go on when this misalignment happens is the occurrence of ethical dissonance. Ethical dissonance can be triggered by the desire to uphold a positive moral self-image, and the temptation and potential benefits associated with unethical behavior (Barkan et al., 2015). The theory describes a conflict between two opposing factors: on the one hand, people want to benefit as much as they can and dishonesty may increase their benefits (Mazar et al., 2008), while on the other hand they want to view themselves as good and honest people (Aronson, 1969; Josephson Institute of Ethics, 2012). Behaving dishonestly may threaten their positive self-concept, but this threat is mediated by the justification of this immoral behavior. These justifications may occur both before (i.e., anticipated ethical dissonance) and after unethical behavior (i.e., experiences ethical dissonance; Shalvi et al., 2015). The empirical evidence is that people are much more prepared to cheat when the extra amount of money or working time is relatively small or can otherwise be rationalized (Ariely, 2012).

Previous research has indicated that social rejection, and the anxiety associated with this rejection, can lead to increased dishonest behavior (Kouchaki and Wareham, 2015). However, not all rejections are social in nature, and the effect of other types of rejection on dishonest behavior remains unclear. In this paper, we focus on the situation in which people's efforts are rejected. We investigate different aspects of the rejection. Specifically, we look into the subjectivity of the rejection and the monetary reward associated with the rejection.

So far, most research on factors that induce dishonest behavior was carried out in the lab. Although lab studies benefit from high experimental control, it remains unclear how findings obtained in a lab translate to an online setting. In a world where technological developments have enabled people to increasingly perform a variety of activities online, it is important to understand how an online context affects people's behavior. Previous dishonesty research has indicated that people may behave differently when they act online. For example, in 15-min long conversations, participants lied more often during online conversations compared to face-to-face interactions (Zimbler and Feldman, 2011). Therefore, we investigate whether the rejection of one's efforts also increases dishonesty in an online setting. This may not only apply to dishonest behavior in general,

but also specifically to the effect of the nature of a rejection on dishonest behavior.

GENERAL MATERIALS AND METHODS

This paper contains three studies that were reviewed and approved by the Research Ethics Committee of Cambridge University's Computer Laboratory. The studies were conducted online. Each experiment started with an information sheet in which participants were informed that they were about to participate in an academic survey (Study 1) or study (Studies 2 and 3). In Study 1, participants were told that the survey was designed to test the language proficiency of the American population. The survey consisted of some general questions and two language related tasks, one grammar and one semantic task. For Studies 2 and 3, participants were told that they were participating in a study to test a newly developed Automatic Validation Tool and that the study involved answering some general questions and filing a mock insurance claim. At this stage, participants were not informed about the true nature of the study, measuring dishonest behavior, because this knowledge could influence their behavior. It was explained that the study would start when clicking "next," and that by doing so they gave their consent to participating in the academic survey/study. At the end of each study, participants were debriefed in writing about the true purpose of the study, and we explained why we could not reveal the deceptive nature of the research earlier. We also asked participants not to share the true nature of the study online until data collection was finished in order to avoid data pollution. As part of the debriefing, all participants were offered the opportunity to contact the experimenters with any questions or complaints, or to retract their data.

The experiments were conducted on Amazon's Mechanical Turk (MTurk), an online platform that is frequently used to collect experimental research data. This recruitment channel was deliberately chosen for two reasons. First, studying dishonest behavior in an anonymous, online environment extends the existing dishonesty literature. Second, experimental research has shown that recruiting on MTurk leads to a representative sample of the U.S. population (Berinsky et al., 2012). This is a more varied participant sample than we would have been able to gather at our university and the variety increases the generalizability of the presented findings within the American population. Another benefit of conducting experimental research on MTurk is the low cost compared to lab experiments, without loss of validity. That low pay does not influence the quality and nature of research results was demonstrated by Paolacci et al. (2010), who replicated a series of classic decision-making studies using MTurk and found similar results to the more expensive original studies that were collected in the lab. Similarly, in Ariely's (2012) experiments, increased financial incentives did not lead to an increase in cheating.

We have studied the effect of different types of rejection (objective, subjective, with promised financial reward, and without) on ethical decision-making using two different types of experimental research designs. We purposefully designed experimental procedures that resembled real-life situations in which the occurrence of dishonest behavior is prevalent. Study 1 comprises a language proficiency study, in which we measured cheating behavior under truly experienced circumstances, while Studies 2 and 3 were both vignette studies in an online insurance claim context that involved participants responding to a hypothetical rather than experienced scenario. Vignette studies have been conducted in a wide range of disciplines including teaching (Poulou, 2001) and nursing (Hughes and Huby, 2002), and have proven particularly useful when studying sensitive topics such as violence in residential care homes (Barter et al., 2004), HIV risk in drug users (Hughes, 1998) and deception (Schweitzer et al., 2004). Due to the sensitive nature of dishonest behavior and insurance fraud, vignettes are a suitable research method for this topic. Although reading a vignette will likely differ from real-world experiences, experimental research has demonstrated that vignettes can provide a sufficiently realistic scenario to affect people's responses (Hughes, 1998; Barter et al., 2004). We additionally added a cover story about testing of our newly developed Automatic Validation Tool to the vignette study to increase the plausibility of our request.

Anxiety, for example when elicited by social rejection, has been shown to affect dishonest behavior (Kouchaki and Desai, 2014; Kouchaki and Wareham, 2015). Because anxiety may also play a mediating role in our rejection manipulation, we invited participants in all three studies to self-report how they felt before and after our manipulation. This allowed for measuring how participants were affected by own rejection manipulation, and whether the elicited emotions mediated the effect of rejection on dishonesty. Additionally, in the first study participants also reported how they felt after the cheating opportunity, to measure if cheating affected how people feel. Dishonesty was measured dichotomously based on actual cheating (Study 1) and lying (Studies 2 and 3) behavior.

STUDY 1

Methods and Materials

Participants and Design

One hundred and sixty-nine American MTurk workers participated in an online study on the effect of unfair rejection on people's mood and cheating behavior. Although the majority of MTurk data is of high quality, some MTurk participants provide random answers. In order to identify these data polluters, we identified several *check questions* in each study. In Study 1, participants had to answer all 10 grammar questions correctly. This conservative criterion was required to operationalize the rejection of effort, see Section "Procedure" in Study 1. Thirteen people failed to answer the 10 questions correctly and were removed from the dataset, leaving 156 participants (94 female; age 18–79, $M = 33.85$, $SD = 13.01$). Of these, three participants did not have English as their native language. Participation took on average 12 min and participants received $1.70 for their time, consisting of a basic payment of $0.50 and a $1.20 bonus that each participant eventually received.

This study is a between-subjects design, measuring the effect of unfair rejection on cheating (cheating vs. no cheating).

Procedure

Participants accessed our website via the MTurk platform and were told that it was a study of English language proficiency, consisting of a grammar test and a semantics test. More specifically, the study was framed as a state-dependent retrieval study (i.e., the memory phenomenon that retrieval performance is affected by the mood and state during which the memories were initially formed; Eich, 1995), in which the effect of mood on test performance would be measured. State-dependent learning was the cover for asking participants to report their feelings regarding five different emotions on a 7-point Likert scale (i.e., happy, sad, guilty, frustrated, and anxious) three times: before and after the feedback (accept or rejection) and after the cheating opportunity. The first two mood questionnaires served a dual purpose. First, they served as a manipulation check, measuring the effect of rejection on participants' feelings. The self-reported emotion ratings serve as a proxy for the perception of the treatment. The second purpose of the first two mood questionnaires was to identify whether the experienced emotions mediate the effect between rejection and cheating. The third questionnaire was included to measure whether the act of cheating affected people's feelings.

The study started with demographic questions, and was subsequently divided into two parts: the grammar and semantics tests. Participants were told that they would receive a $0.60 bonus if they answered all 10 multiple-choice grammar questions correctly. The grammar questions served as a conservative check, and to ensure that participants had invested time and effort before their efforts were rejected. The 10 grammar questions also made our cover story of a language proficiency test more plausible. When participants failed to answer the 10 questions correctly, they received feedback that they would not receive the bonus. As this setting was not a planned manipulation, we excluded these trials from the analysis. Of the participants that answered all 10 questions correctly, half were provided with false feedback that they had answered the final question incorrectly (i.e., rejection condition). The other half did not receive such feedback and were told that they would receive the bonus for this part of the study (i.e., accept condition). Subsequently, participants were asked to provide the definitions of three words. For each correct definition, participants were promised $0.20, with a total of $0.60 if all three definitions were correct. The three words were chosen based on the results of a pilot study in which we investigated what words people do and do not know. Forty-seven were tested, and four participants were removed because they failed to answer the check questions correctly, leaving 43 participants (15 female; age 18–68, $M = 37.21$, $SD = 14.48$). We intended to include two words that all people know, and one word that no one knows. The pilot results indicated that people are familiar with the words 'goal' and 'employee,' but not with the word 'kench.' A kench is a deep bin to salt fish and animal skins, used by fishermen and sailors in the 18 hundreds. Today the word is obsolete and is unknown to the general population.

During Study 1, it was explicitly stated on the website that participants were taking part in a language proficiency test, and that looking up the correct answer was not allowed. Therefore, cheating was defined as providing the correct answer to the 'unknown' target word kench. Twenty-eight participants cheated by providing the correct definition of kench. We also observed another source of unethical behavior, when people quit the experiment with the presumed intention to start over to avoid missing out on the bonus. We had purposefully designed the website such that the back button was disabled and people could only participate from the same IP-address once. Consequently, participants did not succeed when attempting to access the experiment website for a second time. These measures were explicitly explained to all participants and were taken in order to avoid people going back to the previous page to change their answer after receiving the feedback. In total, 13 participants quit after receiving the feedback, and all quitting participants were part of the rejection condition. While participants could quit for other reasons, several participants emailed the experimenter to indicate their intention to start over the study in order to obtain the bonus. Excluding these participants from the dataset would provide a skewed view because it concerns meaningful rather than randomly missing data. Instead of omitting these 13 quitters from the analysis or treating them as cheaters, we consider Behaving Unethically as the broad class of dishonest behaviors, including cheating and quitting. In total, 41 people behaved unethically. After the cheating opportunity, participants were debriefed about the true purpose of the study. All participants received the full bonus of $1.20, regardless of performance and previous feedback. This decision was made in consultation with our ethics committee. This way, all participants were treated equally as payment was not dependent on experimental condition. Because participants were not made aware of this until the data collection was finished, it should not have affected the results.

Results and Discussion

To measure whether our rejection affected people's mood, participants were invited three times to indicate on a 'not at all' (1) to 'very much' (7) Likert scale how happy, sad, anxious, guilty, and frustrated they felt: before (time 1) and after (time 2) the feedback, and after the subsequent cheating opportunity (time 3). Five repeated-measures ANOVAs with Treatment (accept vs. reject) as the independent variable and five self-reported mood measures on times 1 and 2 as the dependent variables revealed four interaction effects, indicating our rejection manipulation was successful. Specifically, participants reported: (i) feeling less happy after being rejected, $F(1,147) = 109.28$, $p < 0.001$, $\eta_p^2 = 0.43$; but, (ii) more sad when rejected, $F(1,147) = 23.69$, $p < 0.001$, $\eta_p^2 = 0.14$; (iii) more frustrated, $F(1,147) = 94.67$, $p < 0.001$, $\eta_p^2 = 0.39$; and, (iv) more anxious, $F(1,147) = 12.94$, $p < 0.001$, $\eta_p^2 = 0.08$. However, guilt was not affected; see **Figure 1**, for a graphical interpretation of the results. Overall, these self-reported mood results indicate that participants' mood was negatively affected by the rejection. To measure whether rejection also

promoted unethical behavior, a chi-square analysis of Treatment on Unethical Behavior was performed. As demonstrated in **Figure 2**, people behave more unethically after rejection (33.3%) compared to being accepted (18.7%), $X^2(2) = 4.32$, $n = 156$, $p = 0.046$, $\Phi = -0.17$. The difference between these two conditions was predominantly caused by the participants that quit the experiment. In the accept condition, 14 out of 75 participants cheated (i.e., 61 participants did not cheat). In the unfair rejection condition, 27 out of 81 participants behaved unethically (i.e., 54 participants did not), of which 14 participants cheated by providing the correct definition of kench and 13 quit early. Because the latter group did not complete the mood questionnaires, it was not possible to run a mediation analysis to determine whether mood mediated the effect between rejection and dishonesty. Finally, to measure whether behaving unethically affected people's emotions, a MANOVA was performed of Unethical Behavior and Treatment on five self-reported mood measures on time 3. This test revealed that people's emotions after the cheating opportunity were affected by Treatment, $F(5,136) = 6.89$, $p < 0.001$, $\eta_p^2 = 0.20$, but not by Unethical Behavior, $F(5,136) = 1.32$, $p = 0.260$, $\eta_p^2 = 0.05$.

A first interesting finding is that cheating in itself did not cause an emotional response. One of the main theories of the detection of deceit is based on the assumption that lying can cause an emotional response (Zuckerman et al., 1981; Vrij, 2008). One hypothetical explanation for this discrepancy between the literature and our finding is that in our study participants cheated (looked up a word while this was not allowed) rather than lied (provide falsified information). Although lying and cheating are both dishonest behaviors, they may elicit different responses. Alternatively, dishonest behavior in general does not always cause an emotional response, for example when there is little at stake. In our study, there were no clear negative consequences to getting caught cheating. If dishonesty does not necessarily elicit an emotion response, this would have consequences for the generalizability of the emotional approach as the base for a lie detection method. This hypothesis has found some support in the deception community, which has shifted from an emotion-based lie detection approach to a cognitive load-based approach over the last decade (Vrij et al., 2015). Future research is needed to investigate the emotional response to different types of dishonest acts in order to determine the generalizability of an emotion-based lie detection approach.

The second interesting finding is that rejection caused both more negative emotions and more dishonest behavior in this online cheating environment. We operationalized rejection by unfairly rejecting the participants' correct answers and thereby taking away their financial reward. That this rejection was perceived as a negative experience was demonstrated by the self-reported mood data; rejection caused people to feel sad, anxious, frustrated, and unhappy. However, based on the current data we cannot determine whether the negative emotional response mediated the effect between rejection and cheating. Therefore, for Study 2, we implemented a new research design in which participants could not cheat by quitting halfway through the experiment. In addition, participants in the current study were rejected on unfair grounds (i.e., although participants answered all questions correctly, they were told they made a mistake and therefore missed out on the financial reward). Therefore, we cannot determine whether rejection in general, or the perceived unfairness of the rejection caused the dishonest behavior. To test whether rejection in general, or the nature of a rejection causes people to behave dishonestly, we conducted a second experiment in which participants were rejected based on objective or subjective criteria.

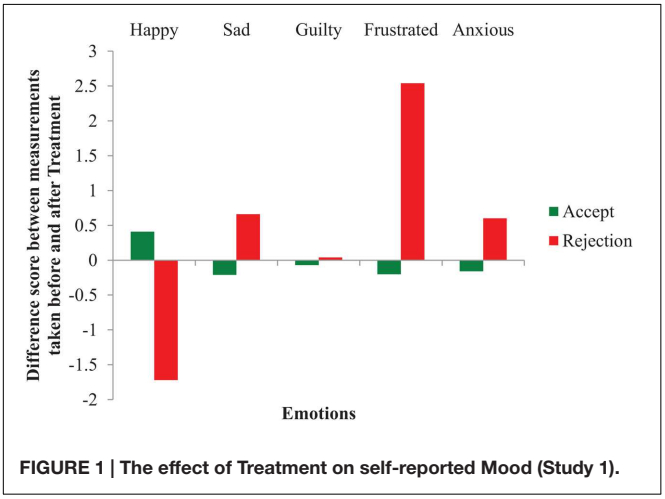

FIGURE 1 | The effect of Treatment on self-reported Mood (Study 1).

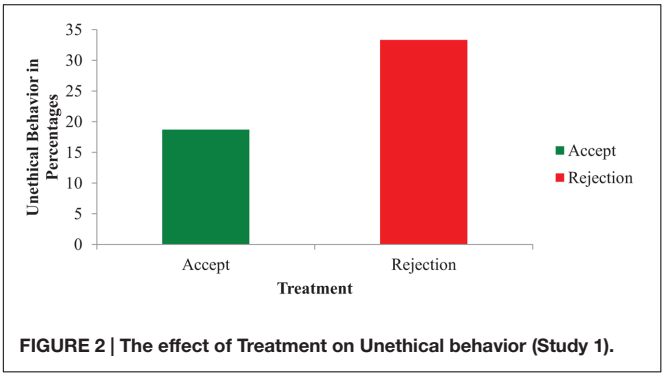

FIGURE 2 | The effect of Treatment on Unethical behavior (Study 1).

STUDY 2

Methods and Materials

Participants and Design

One hundred and forty-four American MTurk workers participated in an online study on the effect of objective and subjective rejections on people's mood and lying behavior. We selected a situation in which dishonesty can occur, and that can leave people with feelings of unfairness: filing an online insurance claim (Derrig, 2002; Topham, 2014). Participants filed a mock insurance claim form. General questions about the trip, such as departure airport and holiday destination, served as check questions. Seven MTurk workers failed to answer these questions correctly and were removed from the dataset, leaving

137 participants (75 female; age 19–72, $M = 35.15$, $SD = 11.25$). Participation took on average 14 min and participants received $1.00 for their time, consisting of a basic payment of $0.50 and a $0.50 bonus that every participant eventually received.

This study is a between-subjects design, measuring the effect of verdict (accept, objective reject vs. subjective reject) on falsifying information (falsified vs. not falsified information).

Procedure

Participants accessed our website via the MTurk platform and were told that this study was designed to test the accuracy and usability of an Automatic Validation Tool for online insurance claims. We asked participants to imagine that their backpack was stolen during a holiday trip in Europe, and they were asked to file a mock insurance claim form for this stolen backpack. To ensure consistency between sessions, participants were provided with a scenario describing how their backpack was stolen, and they received an overview of the main guidelines of 'their' travel insurance. This information was accessible through a pop-up menu whilst completing the insurance claim to ensure that participants would not make mistakes due to memory errors, rather than the deliberate falsification of information. To mimic a real-world situation, participants were promised a monetary reward (i.e., $0.50 bonus) upon claim acceptance. Both the insurance claim form and the policy guidelines were based on information provided by a large UK-based insurance company. The scenario, policy guidelines and claim form can be found at https://www.projects.science.uu.nl/lyingfeelsright/. Participants were also asked to complete two mood questionnaires, once before filing the claim and once after hearing the verdict on their claim, followed by the lie opportunity during which participants could falsify information on their insurance claim in order to get the claim accepted. Participants were led to believe this mood questionnaire to be part of testing the usability of the Automatic Validation Tool, while it actually was aimed at measuring whether people's feelings were affected by a rejected claim.

The study started with demographic questions, followed by the presentation of the scenario and policy guidelines. Subsequently, participants filed an insurance claim based on the backpack scenario. After submitting their claim, participants saw the following message for several seconds: "Please wait a moment. We are now automatically checking the content of your insurance claim. Do not push the back button or refresh." This message was followed by the verdict on their claim. After the verdict, people had the possibility to complain and/or make changes to their claim if they believed the Automatic Validation Tool had misunderstood what happened. Falsifying information (i.e., behaving dishonestly) was defined as submitting information that diverged from the information presented in the scenario, which could help participants get their claim accepted and win a reward.

Participants were randomly assigned to one of three Verdict conditions: accept, objective rejection and subjective rejection. In the accept condition, participants were told that, based on information they provided, their claim got accepted and that they would receive the bonus. In the objective rejection condition, participants were told that their claim was rejected because they had violated the maximum journey limit of 31 days. The scenario in all conditions was identical, except for the length of the journey in the objective rejection, which was 5 weeks instead of three, exceeding the insurance policy journey limit. We designed the violation of maximum journey length because it concerns a clearly-stated, common policy guideline, and one familiar to the general public. It can therefore be regarded as an objective rejection. Although this rejection was based on objective policy guidelines, participants still had the opportunity to cheat because they could change their departure or return date, or mention that they notified the insurance company of their extended journey beforehand.

The subjective rejection was based on the wide interpretation of the ambiguous statement that "people should take care to look after their personal possessions, in particular their valuables." Participants were told that their claim was rejected because, based on the provided information, the conclusion had been drawn that they had been negligent in taking care of their possessions. There is no clear description of what behaviors do and do not count as negligence, making this a subjective rejection. Participants could falsify information by fabricating more convincing ways (i.e., not described in the provided scenario) in which they had taken care of their backpack. For example, one participant claimed that he had not left the backpack out of his sight, while it was clearly stated in the scenario that he/she only realized the backpack was missing when he/she was about to leave the restaurant. For each participant, a coder determined whether any information in the statement contradicted information provided in the scenario. For the objective rejection condition, this included mentioning incorrect dates and prior contact with the insurance company. For the subjective rejection, this included mentioning contact with the thief, keeping the backpack in sight at all times, and incorrect information about the location of their backpack. Other statements that participants made to increase the chances of getting their claim accepted, but which did not contradict information from the scenario such as claims of being an honest and careful person and portraying feelings of unfair treatment were not interpreted as false information. After the lie opportunity, participants were debriefed about the true purpose of the study, and all participants received the $0.50 bonus.

Results and Discussion

In line with Study 1, to measure participants' emotions, five repeated-measures ANOVAs were performed with Verdict (i.e., accept, objective reject, and subjective reject) as the independent variable and the five self-reported Mood questions as the dependent variables. Results indicated with five interaction effects that participants' mood was negatively affected by rejection in general, and that the nature of the rejection did not matter. Specifically, participants reported: (i) feeling less happy when getting rejected in general, compared to getting their claim accepted, $F(2,134) = 73.13$, $p < 0.001$, $\eta_p^2 = 0.52$; but, (ii) more sad when getting rejected in general, compared to getting their claim accepted, $F(2,134) = 21.78$, $p < 0.001$, $\eta_p^2 = 0.25$; (iii) more guilty when the rejection was subjective, compared to an objective rejection and acceptance, $F(2,134) = 7.34$, $p = 0.001$, $\eta_p^2 = 0.10$; (iv) more frustrated when rejected in general, compared to

getting accepted, $F(2,134) = 58.43$, $p < 0.001$, $\eta_p^2 = 0.47$; and, (v) more anxious when rejected in general, compared to getting accepted, $F(2,134) = 4.17$, $p = 0.018$, $\eta_p^2 = 0.02$. See **Figure 3**, for a graphical interpretation of the results. These self-reported results indicate that, overall, participants' mood was negatively affected by claim rejection, regardless of the nature of this rejection. Feelings of guilt were the only exception, as participants who were subjectively rejected felt guiltier after this than participants who were rejected based on measurable criteria or who were not rejected at all. A follow-up correlational analysis of Falsified information on self-reported Guilt in participants in the subjective rejection condition revealed that guilt was not induced by the dishonest behavior (i.e., falsifying information on the insurance claim form), $r = 0.03$, $n = 47$, $p = 0.861$.

To measure whether rejection promoted dishonest behavior, a chi-square analysis of Verdict on the dependent variable Falsified information (i.e., yes or no) was performed. As demonstrated in **Figure 4**, people lie more after being rejected based on subjective criteria (23.4%) compared to rejection based on objective criteria (14.9%), and getting accepted (0%), $X^2(2) = 10.97$, $n = 137$, $p = 0.004$, $\Phi = 0.28$. To identify whether rejection in general caused this effect, or whether the nature of the rejection played a role as well, we ran an additional chi-square analysis in which we removed the accept condition. Results demonstrate that although participants falsified information more often when rejected subjectively by 8.5%, this difference between the two reject conditions was not significant, $X^2(1) = 1.09$, $n = 94$, $p = 0.294$, $\Phi = 0.11$. We next conducted a multiple mediation analysis following procedures by Preacher and Hayes (2008) to test whether self-reported mood (i.e., happiness, sadness, guilt, frustration, and anxiety) mediates the effect of rejection on dishonest behavior. We ran a bootstrapping analysis (5000 iterations) with the five mood variables simultaneously in the model and results indicated that only happiness [0.081 1.003] mediated the effect between rejection and dishonest behavior. The 95% bias corrected confidence intervals of sadness, guilt, frustration, and anxiety included zero, suggesting that these variables did not have a mediating effect.

In summary, rejection in general (i.e., regardless of the nature of this rejection), leads to negative emotions and more dishonest behavior. To which extent the nature of a rejection increases dishonest behavior, is a topic for further research. Participants falsified information more often and, independently, experienced more feelings of guilt when the rejection was based on subjective reasons. However, the chi-square analysis did not support this finding. Whether the difference in dishonesty between objective and subjective rejections was not significant due to a lack of power, or because dishonesty is predominantly driven by rejection rather than the nature of the rejection, cannot be determined based on the current data.

In the previous two studies, rejection always led to financial loss. Participants were also aware that they would profit financially from cheating (i.e., looking up the correct definition) and lying (i.e., falsifying information on an insurance claim to increase the chance of claim acceptance). Therefore, based on the results from Studies 1 and 2, we cannot disentangle whether people behaved dishonestly to compensate for their previous financial loss, or due to the rejection and negative emotions elicited by these rejections. We wondered, would this behavior change if there were no financial incentives associated with dishonesty? In other words, are people just trying to get back the money that was unfairly taken for them, or are they emotionally seeking revenge?

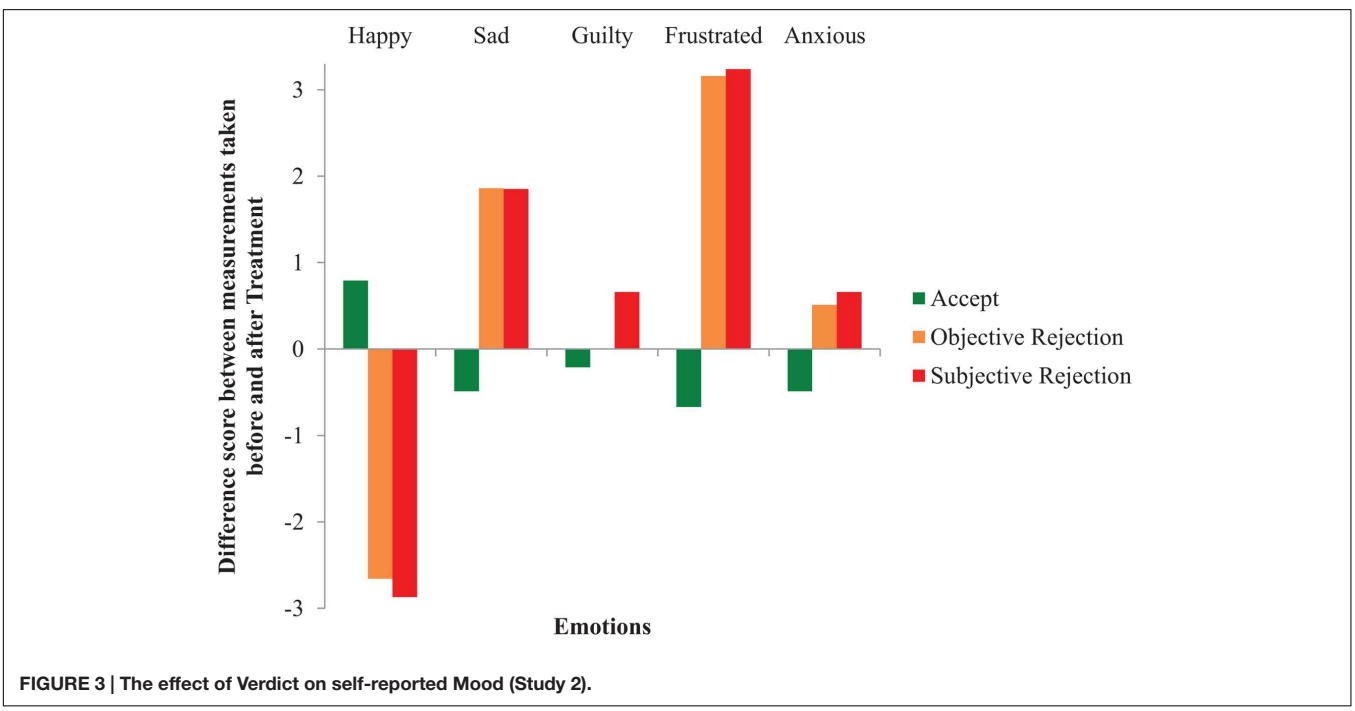

FIGURE 3 | The effect of Verdict on self-reported Mood (Study 2).

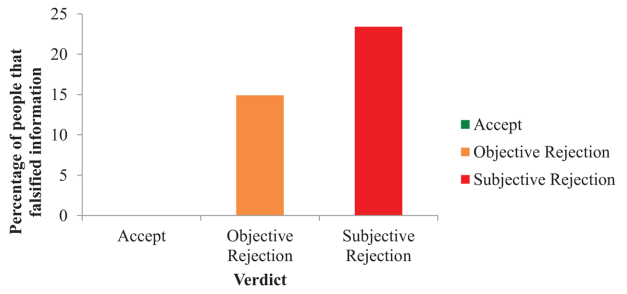

FIGURE 4 | The effect of Verdict on the percentage of participants that falsified information (i.e., lying; Study 2).

STUDY 3

In the third and final experiment we tackled the confounding effect that dishonesty will often lead to financial gain (Greenberg, 1993; Houser et al., 2012; Kline et al., 2014); we removed the financial incentive to cheat in order to investigate whether financial incentives are the main motivator for people's behavior when they have been rejected, based on either objective or subjective criteria.

Methods and Materials

Participants and Design

One hundred and seventy-nine American MTurk workers participated in an online study on the effect of rejection and monetary rewards on people's mood and lying behavior. Three participants failed to answer the check questions that were based on general scenario information correctly and were removed from the dataset, leaving 176 participants (110 female; age 18–68, $M = 36.81$, $SD = 12.29$). Participation took on average 14 min and participants received $1.00 for their time, consisting of a basic payment of $0.50 and a $0.50 bonus that every participant received.

This study is a 3 × 2 between-subjects design, measuring the effect of verdict (accept, objective reject vs. subjective reject) and bonus (bonus vs. bonus-after) on falsifying information (falsified vs. not falsified information).

Procedure

The procedure of Study 3 follows the procedure of Study 2 with one exception. Instead of telling all participants at the beginning of the experiment that they would receive a $0.50 bonus upon claim acceptance, half of the participants were not told about the bonus until the debriefing. In other words, participants in the bonus-after condition were not promised any financial bonus during the experiment and were only made aware of the existence of the bonus upon completion of the experiment. This way, we could measure the effect of a prospective bonus on mood and lying behavior. After the lie opportunity, participants were debriefed about the true purpose of the study, and all participants, including the participants in the bonus-after condition, received the $0.50 bonus at the end of the study.

Results and Discussion

To measure if our Bonus manipulation affected participants' emotions, five repeated-measures ANOVAs were conducted with Verdict (i.e., accept, objective reject, and subjective reject), and Bonus (i.e., bonus and bonus-after), as the independent variables and five self-reported Mood questions as the dependent variables. The mood results from Study 2 were replicated with five interaction effects, indicating that participants' mood was negatively affected by getting a claim rejected in general, regardless of the nature of a rejection. Specifically, participants reported: (i) feeling less happy when getting rejected in general, compared to accepted, $F(2,170) = 84.57$, $p < 0.001$, $\eta_p^2 = 0.50$; but, (ii) more sad when getting their claim rejected in general, compared to accepted, $F(2,170) = 33.61$, $p < 0.001$, $\eta_p^2 = 0.28$; (iii) more guilty when rejected in general, $F(2,170) = 4.89$, $p = 0.009$, $\eta_p^2 = 0.05$; (iv) more frustrated when rejected in general, $F(2,170) = 49.40$, $p < 0.001$, $\eta_p^2 = 0.37$; and, (v) more anxious when rejected in general, $F(2,170) = 20.48$, $p < 0.001$, $\eta_p^2 = 0.19$. See **Figures 5–7** for graphical interpretations of these results. Importantly, in addition to Verdict, Bonus also affected

FIGURE 5 | The effect of Bonus on self-reported Mood in the Accept condition (Study 3).

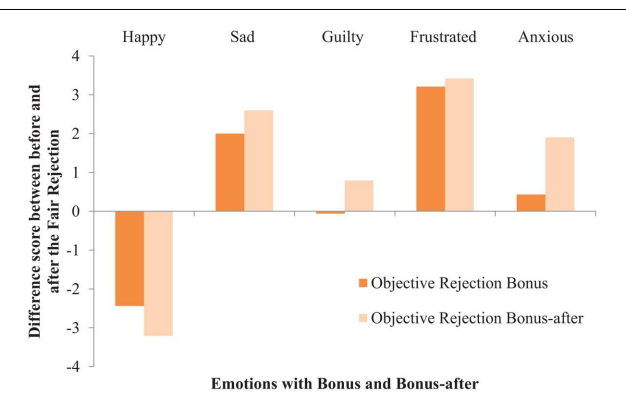

FIGURE 6 | The effect of Bonus on self-reported Mood in the Objective rejection condition (Study 3).

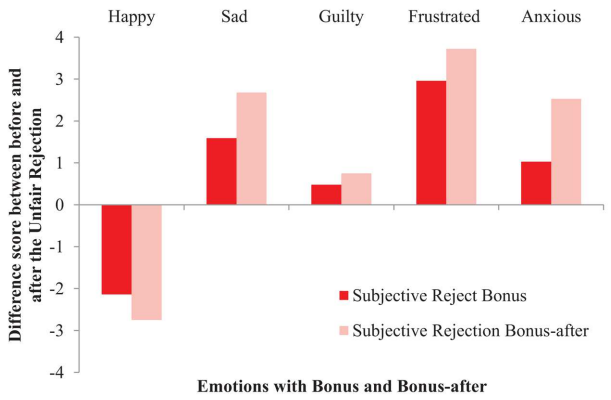

FIGURE 7 | The effect of Bonus on self-reported Mood in the Subjective Rejection condition (Study 3).

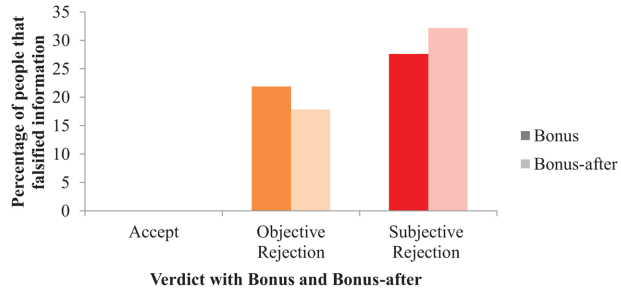

FIGURE 8 | The effect of Verdict and Bonus on the percentage of participants that falsified information (i.e., lying; Study 3).

people's mood. Participants reported feeling less happy after hearing the verdict about their claim when they were not aware of the bonus ($M = 3.37$, $SD = 2.22$), compared to situations in which participants received a bonus upon claim acceptance ($M = 3.78$, $SD = 2.20$), $F(1,170) = 4.57$, $p = 0.034$, $\eta_p^2 = 0.03$. Participants also reported feeling more guilty after hearing the verdict in the bonus-after condition ($M = 2.00$, $SD = 1.48$), compared to scenarios where participants received a bonus, ($M = 1.62$, $SD = 1.15$), $F(1,170) = 4.85$, $p = 0.029$, $\eta_p^2 = 0.03$. Lastly, participants reported feeling more anxious after hearing the verdict in the bonus-after condition ($M = 3.23$, $SD = 2.13$), compared to the bonus condition ($M = 2.49$, $SD = 1.99$), $F(1,170) = 14.52$, $p < 0.001$, $\eta_p^2 = 0.08$. These self-reported mood results indicate that, overall, participants' mood was negatively affected when their claim was rejected, regardless of the nature of this rejection. These results also indicate that people felt emotions more strongly when there was seemingly no money involved (i.e., decreased happiness, increased guilt and anxiety).

To measure whether a monetary bonus not only affected people's mood, but also their tendency to behave dishonestly, a loglinear analysis of Verdict (i.e., accept, objective reject, and subjective reject), and Bonus (i.e., bonus and bonus-after) on Falsified information (i.e., yes or no) was performed. In other words, we analyzed if people's tendency to lie was dependent on monetary rewards and the nature of their claim rejection. In line with standard practice, 0.5 was added to all cells to avoid performing calculations with empty cells. The loglinear regression revealed that the highest order three-way model did not retain all effects. Instead, the best fit was a second-order model, $X^2(0) = 0$, $p = 1$, including a two-way interaction effect between Verdict and Falsified information, $X^2(5) = 28.08$, $n = 176$, $p < 0.001$. A separate chi-square analysis of Verdict on Falsified information demonstrated a significant difference between the conditions unfairly rejected (29.8%), fairly rejected (20.0%), and accepted (0%), $X^2(2) = 19.56$, $n = 176$, $p < 0.001$, $\Phi = 0.33$, see **Figure 8**. Although participants in the subjective reject condition cheated almost 10% more (and relatively 49% more) than participants in the objective reject condition, this effect did not differ significantly when tested without the accept condition, $X^2(1) = 1.51$, $n = 117$, $p = 0.219$, $\Phi = 0.11$. The influence of Bonus on Falsified Information was not further tested because these factors were not included in the best fitting loglinear regression model. We next conducted two multiple mediation analyses to test whether self-reported mood (i.e., happiness, sadness, guilt, frustration, and anxiety) mediates the effect of rejection on dishonest behavior. Because the presence or seemingly absence of a bonus influenced how rejection affected people's mood, we split the file up based on Bonus condition and ran two separate analyses. We ran two bootstrapping analyses, one for the bonus and one for the bonus-after condition (5000 iterations each) with the five mood variables simultaneously in the model. Results indicated that none of the emotions mediated the effect of rejection on subsequent dishonest behavior.

In conclusion, the mood results demonstrate that people experienced the negative emotions associated with rejection more strongly when there was no financial reward involved, although these strong emotions did not subsequently increase dishonest behavior. Dishonesty was also not affected by the presence of a financial reward, or by the nature of the rejection. Instead, rejections in general fueled dishonest behavior. So the absence of money made people care more, but it did not spark dishonesty: being rejected did.

GENERAL DISCUSSION

In previous research, dishonesty was often quantified as cheating (Houser et al., 2012) or stealing (Greenberg, 1993). Because dishonesty encompasses more behaviors than just stealing and cheating, dishonest behavior in this paper was not only quantified as cheating (i.e., on a test; Study 1), but also as falsifying information (i.e., on a mock insurance claim; Studies 2 and 3). These types of dishonest behaviors are, for example, relevant in the applied context of insurance claims. Because insurance claims are nowadays often filed online, the studies in this paper were conducted using the online platform Mechanical Turk.

In three studies, we have investigated whether the rejection of one's efforts elicits dishonest behavior in an online setting. In Study 1, we pretended to run a language proficiency study and rejected the efforts of half of the participants by providing

them with negative feedback about their performance. People who were rejected subsequently engaged more often in unethical behavior when they got the chance than people who did not previously got rejected. Although this study provided an interesting insight in online dishonest behavior, we discovered two possibly confounding factors. First, because participants in the reject condition were rejected unfairly (i.e., although they had answered all questions correctly, they were told they made a mistake), we could not determine whether the rejection in itself, or the nature of the rejection caused the rise in dishonest behavior. Second, the rejection resulted in financial loss, which meant we could not determine whether the rejection, or the financial loss associated with the rejection caused people to behave dishonestly. In Study 2, we tested whether the nature of the rejection, operationalized as rejection based on objective or subjective criteria, affected dishonest behavior. In Study 3, we investigated the role of financial rewards in the motivations for dishonest behavior. These two studies were conducted in an online insurance claim environment for its relevance in our current society. Results indicated that across experiments, the rejection, rather than the nature of a rejection, promoted dishonest behavior. Although participants cheated approximately 10% more (a relative difference of 49%) after being rejected for subjective reasons compared to objective criteria, this difference was not statistically significant. Based on the current data we cannot determine whether the nature of a rejection simply does not affect dishonest behavior, or whether the lack of effect was due to a lack in power. Other papers investigating the effect of treatment on dishonesty experienced similar power problems. For example, the difference in dishonest behavior between the fair and unfair condition in Houser et al.'s (2012) study was only marginally significant with a sample size of 500+. This suggests that the effect size of fairness on dishonesty may be relatively small. Regardless, decreasing dishonest behavior in the context of insurance claims with a few percent can still lead to large financial benefits, making this topic worth exploring.

When removing the financial rewards associated with dishonest behavior, participants still made an effort to falsify information, suggesting that dishonesty is not just caused by an attempt to get restorative justice for missing out. Although having a financial reward associated with accepted claims – as is common in real-life insurance claims – affected people's feelings, it did not affect dishonest behavior. These results support previous theory (Ariely, 2012) and experimental results (Mazar et al., 2008) on the irrationality of dishonesty, which demonstrates that people do not base their decision to behave dishonestly on a rational cost-benefit analysis. We tested this by adding the bonus-after condition in Study 3, so in the perception of the participants we removed the (financial) benefits of acting dishonestly, while the costs in terms of effort did not change. If people were rational economic actors, the seemingly absence of financial rewards would have stopped them from cheating. However, dishonest behavior was not affected. Rather, when there are no financial gains in prospect, emotional involvement was larger. It is as if playing for honor is more important than playing for money. When unaware of any prospective reward, participants indicated feeling less happy, and more guilty and anxious after hearing the verdict about their claim than people who had hoped for financial benefits. Importantly, although the financial benefits in Study 3 were small, they still elicited emotional and behavioral changes. Larger incentives would not necessarily have increased this effect, just as Ariely (2012) demonstrated that increasing the financial incentive did not lead to increased cheating. Moreover, Ruedy et al. (2013) replaced the financial incentive with a more personal incentive to cheat by linking success on the test to intelligence and professional success in life and found that people cheated significantly more when their self-esteem was at stake.

A theory that may help explain these irrational dishonesty results is 'ethical dissonance' (Barkan et al., 2015), a theory that is related to the general 'cognitive dissonance' theory by Festinger (1957) in which internal consistency is threatened by two or more conflicting beliefs and ideas. Specifically, in our insurance claim studies, ethical dissonance may have occurred when people tried to justify to themselves why they spent time and effort (i.e., adding feedback and falsifying information on the claim form) without any potential benefits (i.e., no monetary bonus). The friction caused by these conflicting beliefs may then be solved by stating that they made this effort because they care (i.e., higher emotional involvement). The ethical dissonance theory (Barkan et al., 2015) describes how people feel torn between wanting to be a good person (Aronson, 1969; Josephson Institute of Ethics, 2012), and wanting the benefits of behaving dishonestly (Mazar et al., 2008). Although we did not directly ask participants how they felt about themselves in order to avoid priming (dis)honest behavior (Mazar et al., 2008; Shu et al., 2012), the implemented mood questionnaires can be used as an indication of their mental states. In previous research, dishonest behavior has been linked both to eliciting negative emotions such as guilt (Massi, 2005) and to eliciting positive affect (Ruedy et al., 2013). Here, the mood results from the first study showed that cheating did not affect people's emotions, suggesting that our participants may have been effective at justifying their dishonest behavior. This is key, because dishonesty can be a slippery slope (Lerman, 2002; Ariely, 2012). If people can behave dishonestly and still feel good or even better (Ruedy et al., 2013) about themselves, they might be more likely to behave dishonestly again in the future.

The chosen experimental designs have several benefits including the ability to test multiple types of dishonest behavior. This allowed us to investigate aspects of dishonesty that go beyond simple tasks such as reporting the outcome of a coin toss (Houser et al., 2012), and analyze dishonest behavior in more realistic settings. However, when participants complete a study on their own computers, this typically reduces the amount of experimental control. Specifically in Study 1, we could not distinguish between participants who quit in an attempt to cheat, and those who quit for other reasons such as frustration or lack of trust in the system. In Studies 2 and 3, participants might have reported more negative emotions, not as the sole result of the feedback decision, but caused by a discrepancy between behavior dictated by their assignment, and the behavior that would have led to the highest gain. More specifically, when participants in the objective rejection condition filled out the travel dates conscientiously, their claim would get rejected. While the online

insurance fraud scenario provides both realism and a structured experimental testing mechanism for dishonest behavior, the scenarios and instructions may have posed conflicting incentives for the participants. In addition, much dishonesty research shows that people usually cheat and lie a little (Ariely, 2012; Houser et al., 2012). The complete absence of falsifying information in the accept conditions of Studies 2 and 3 is likely to have been caused by the choice of experimental design because claim acceptance, and therefore pay-off, was quantified as a binary decision. In other words, an accepted claim led to the highest achievable monetary reward and therefore did not require participants to falsify additional details, whilst in real life people could still inflate their claim a little, and thus receive more money.

The consistency of our dishonesty findings across three studies and two research designs strengthens our belief that people behave more dishonestly after rejection, specifically in an online environment. In an applied setting, this would imply that firms should try to minimize the amount of rejected claims, for example through heightened transparency and clearer communication of acceptance guidelines. When it is upfront clear whether a claim is likely to be accepted or not, people may submit less claims that clearly violate policy guidelines, leading to a reduction in rejection and thus subsequent dishonest behavior. Despite the lack of a statistically significant difference, likelihood ratios indicated that the nature of a rejection may contribute to the elicitation of dishonest behavior as well. Because even a small decrease in fraudulent insurance claims can lead to a large savings, stating the rejection policy more clearly could not only reduce the amount of rejected claims, it may also further reduce dishonest behavior when people feel that they were rejected on objective grounds.

Although rejection did cause people to feel more negative (i.e., less happy, more sad, frustrated, and anxious), this emotional response did not have a strong effect on dishonest behavior. The mediation analyses of Studies 2 and 3 indicated that the majority of emotions did not affect dishonesty as a mediator (i.e., indirectly). Only the reduction in happiness in Study 2 mediated the effect between rejection and dishonesty. Anxiety, a previously demonstrated mediator in the context of social rejection (Kouchaki and Wareham, 2015), did not have a similar effect in our studies. There are several hypothetical explanations for the discrepancy between our findings and the existing literature on this topic. First, social rejections may elicit a different response than rejected efforts. The relationship between social rejection and (social) anxiety is well explored and lies at the core of human functioning (Baumeister and Tice, 1990; Leary, 1990). The rejection of one's efforts may play a less central role and therefore have a weaker corresponding anxiety effect. A second possible explanation could be that people behave differently online, compared to face-to-face situations. The majority of research on factors that affect dishonest behavior has been conducted in lab experiments, but the few studies that have investigated dishonesty in an online context have demonstrated that the extent to which people behave dishonestly is affected by the modality of their interaction. For example, Zimbler and Feldman (2011) found that people tend behave more dishonestly when interacting online.

A factor that may have mediated the effect between rejection and dishonesty, but which we did not explicitly test, is fairness. In Studies 2 and 3, we differentiated between rejections based on objective and subjective grounds. Especially the rejections on subjective grounds may have elicited feelings of unfairness. Previous research has demonstrated that fairness can induce dishonest feelings (e.g., satisfaction levels; Hegtvedt and Killian, 1999), plans (e.g., hypothetical dishonest behavior; Schweitzer and Gibson, 2008), and even behavior (e.g., selfish behavior, Kline et al., 2014; cheating, Houser et al., 2012; and stealing money, Greenberg, 1993). Fairness has also proven be to an influential factor when it comes to online behavior, as the fairness of a request was the best predictor of honest behavior in a personal information disclosure study (Malheiros et al., 2013). Whether violations of fairness mediate the effect between rejection and dishonesty, will need to be explored in future research. If fairness turns out to be influential, firms can further experiment with attempting to adapt their customers' fairness perceptions. The fairness of a situation is often ambiguous (Van den Bos et al., 1997), and fairness perceptions can be influenced (Bies and Shapiro, 1987; Egelman et al., 2010). In other words, violations of fairness principles may be used to justify dishonest behavior that is ambiguous, a factor that has repeatedly been shown to justify dishonesty (Schweitzer and Hsee, 2002; Shalvi et al., 2015). Therefore, investigating what factors determine whether customers interpret a situation as fair will allow firms to promote honest behavior by tipping the conflict between wanting to be a good person and the benefits of dishonesty in the honest direction (Bies and Shapiro, 1987). Transparency may be the way for firms to see to it that their customers do not feel that lying is the right thing to do, potentially reducing the cost of opportunistic fraud.

AUTHOR CONTRIBUTIONS

All authors helped design the experiments and commented on the paper. SV collected the data and wrote the initial draft. RP created the framework for the online data collection. All co-authors revised the paper for resubmission.

FUNDING

This research was performed as part of the UK Engineering and Physical Sciences Research Council.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Lara Warmelink for providing useful feedback. We also thank the reviewers for their constructive comments that helped to improve the paper.

REFERENCES

Alanezi, F., and Brooks, L. (2014). "Combatting online fraud in Saudi Arabia using general deterrence theory," in *Proceeding of the 20th Americas Conference on Information Systems,* Savannah, GA.

Anderson, R., Barton, C., Böhme, R., Clayton, R., van Eeten, M. J. G., and Levi, M. (2012). "The economics of information security and privacy," in *Measuring the Cost of Cybercrime,* eds R. Böhme (Berlin: Springer)

Ariely, D. (2012). *The Honest Truth about Dishonesty: How we lie to Everyone - Especially Ourselves*. New York, NY: Harper Collins.

Aronson, E. (1969). "The theory of cognitive dissonance: a current perspective," in *Advances in Experimental Social Psychology*, ed. L. Berkowitz (New York, NY: Academic Press), 1–34.

Barkan, R., Ayal, S., and Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Curr. Opin. Psychol.* 6, 157–161. doi: 10.1016/j.copsyc.2015.08.001

Barter, C., Renold, E., Berridge, D., and Cawson, P. (2004). *Peer Violence in Children's Residential Care*. Basingstoke: Palgrave Macmillan.

Baumeister, R. F., and Tice, D. M. (1990). Point-counterpoints: anxiety and social exclusion. *J. Soc. Clin. Psychol.* 9, 165–195. doi: 10.1521/jscp.1990.9.2.165

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Polit. Anal.* 20, 351–368. doi: 10.1093/pan/mpr057

Bies, R. J., and Shapiro, D. L. (1987). Interactional fairness judgments: the influence of causal accounts. *Soc. Justice Res.* 1, 199–218. doi: 10.1007/BF01048016

Blasi, A. (1980). Bridging moral cognition and moral action: a critical review of the literature. *Psychol. Bull.* 88, 1–45. doi: 10.1007/s10979-009-9185-9

Button, M., Lewis, C., and Tapley, J. (2014). Not a victimless crime: the impact of fraud on individual victims and their families. *Security J.* 27, 36–54. doi: 10.1057/sj.2012.11

Centre for Counter Fraud Studies (2014). *Counter Fraud 2020: A Twelve Point Plan to Protect the UK*. Portsmouth: Centre for Counter Fraud Studies.

Cross, C. (2013). "Nobody's holding a gun to your head. . ." examining current discourses surrounding victims of online fraud crime, justice and social democracy," in *Proceedings of the 2nd International Conference, Crime and Justice Research Centre,* eds R. Kelly and T. Juan (Brisbane, QLD: Queensland University of Technology), 25–32.

Derrig, R. A. (2002). Insurance fraud. *J. Risk Insur.* 69, 271–287. doi: 10.1111/1539-6975.00026

Egelman, S., Molnar, D., Christin, N., Acquisti, A., Herley, C., and Krishnamurthi, S. (2010). "Please continue to hold: an empirical study on user tolerance of security delays," in *Proceeding of the 9th Workshop on Economics of Information Security*, Cambridge, MA.

Eich, E. (1995). Searching for mood dependent memory. *Psychol. Sci.* 6, 67–75. doi: 10.1111/j.1467-9280.1995.tb00309.x

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Gee, J., and Button, M. (2013). *The Financial Cost of Fraud Report 2013*. Portsmouth: BDO LLP. Available at: http://www.bdo.co.uk/_data/assets/pdf_file/0004/16942/The-Financial-Cost-of-Fraud.pdf

Gino, F., and Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *J. Pers. Soc. Psychol.* 102, 445–459. doi: 10.1037/a0026406

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behaviour: the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gino, F., Norton, M. I., and Ariely, D. (2010). The counterfeit self: the deceptive costs of faking it. *Psychol. Sci.* 21, 712–720. doi: 10.1177/0956797610366545

Greenberg, J. (1993). Stealing in the name of justice: informational and interpersonal moderators of theft reactions to underpayment inequity. *Organ. Behav. Hum. Decis. Process.* 54, 81–103. doi: 10.1006/obhd.1993.1004

Hegtvedt, K. A., and Killian, C. (1999). Fairness and emotions: reactions to the process and outcomes of negotiations. *Soc. Forces* 78, 269–302. doi: 10.2307/3005797

Houser, D., Vetter, S., and Winter, J. (2012). Fairness and cheating. *Eur. Econ. Rev.* 56, 1645–1655. doi: 10.1016/j.euroecorev.2012.08.001

Hughes, R. (1998). Considering the vignette techniques and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociol. Health Illness* 20, 381–400. doi: 10.1006/obhd.1993.1004

Hughes, R., and Huby, M. (2002). The application of vignettes in social and nursing research. *J. Adv. Nurs.* 37, 382–386. doi: 10.1046/j.1365-2648.2002.02100.x

Josephson Institute of Ethics (2012). *2012 Report card on the ethics of American youth*. Available at: http:charactercounts.org/programs/reportcard [accessed on 12 September, 2014].

Kline, R., Galeotti, F., and Orsini, R. (2014). "When foul play seems fair: dishonesty as a response to violations of just deserts," in *Proceedings of the Quaderni – Working Paper DSE (920)*, Department of Economics, University of Bologna, Bologna.

Kouchaki, M., and Desai, S. D. (2014). Anxious, threatened, and also unethical: how anxiety makes individuals feel threatened, and commit unethical acts. *J. Appl.* 100, 360–375. doi: 10.1037/a0037796

Kouchaki, M., and Smith, I. H. (2014). The morning morality effect: the influence of time of day on unethical behavior. *Psychol. Sci.* 25, 95–102. doi: 10.1177/0956797613498099

Kouchaki, M., and Wareham, J. (2015). Excluded and behaving unethically: social exclusion, physiological responses, and unethical behavior. *J. Appl. Psychol.* 100, 547–556. doi: 10.1037/a0038034

Leary, M. R. (1990). Responses to social exclusion: social anxiety, jealousy, loneliness, depression, and low self-esteem. *J. Soc. Clin. Psychol.* 9, 221–229. doi: 10.1521/jscp.1990.9.2.221

Lerman, L. G. (2002). The slippery slope from ambition to greed to dishonesty: lawyers, money and professional integrity. *Hofstra Law Rev.* 30, 1–44.

Malheiros, M., Preibusch, S., and Sasse, M. A. (2013). "Fairly truthful": the impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. *Trust Trustworthy Comput.* 7904, 250–266. doi: 10.1007/978-3-642-38908-5_19

Massi, L. L. (2005). Anticipated guilt as behavioral motivation. *Hum. Commun. Res.* 31, 453–481. doi: 10.1111/j.1468-2958.2005.tb00879.x

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Market.* 45, 633–653.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgm. Decis. Mak.* 5, 411–419.

Poulou, M. (2001). The role of vignettes in the research of emotional and behavioural difficulties. *Emot. Behav. Difficult.* 6, 50–62. doi: 10.1080/13632750100507655

Preacher, K. J., and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891. doi: 10.3758/BRM.40.3.879

Purkait, S. (2012). Phishing counter measures and their effectiveness: literature review. *Inform. Manag. Comput. Security* 20, 382–420. doi: 10.1108/09685221211286548

Ruedy, N. E., Moore, C., Gino, F., and Schweitzer, M. (2013). The cheater's high: the unexpected affective benefits of unethical behavior. *J. Pers. Soc. Psychol.* 105, 531–548. doi: 10.1037/a0034231

Schweitzer, M., and Gibson, D. (2008). Fairness, feelings, and ethical decision making: consequences of violating community standards of fairness. *J. Bus. Ethics* 77, 287–301. doi: 10.1007/s10551-007-9350-3

Schweitzer, M., and Hsee, C. K. (2002). Stretching the truth: elastic justification and motivated communication of uncertain information. *J. Risk Uncertainty* 25, 185–201. doi: 10.1023/A:1020647814263

Schweitzer, M., Ordóñez, L., and Douma, B. (2004). Goal setting as a motivator of unethical behavior. *Acad. Manag. J.* 47, 422–432. doi: 10.2307/20159591

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications: doing wrong and feeling moral. *Curr. Direct. Psychol. Sci.* 24, 125–130. doi: 10.1177/0963721414553264

Shu, L. L., Mazar, N., Gino, F., Ariely, D., and Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15197–15200. doi: 10.1073/pnas.1209746109

Topham, G. (2014). *Ryanair and EasyJet fined a total €1m in Italy for mis-selling travel insurance*. Available at: http://www.theguardian.com/business/2014/feb/17/ryanair-easyjet-fined-travel-insurance [accessed on 12 September, 2014].

Van den Bos, K., Lind, E. A., Vermunt, R., and Wilke, H. A. M. (1997). How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. *J. Pers. Soc. Psychol.* 72, 1034–1046. doi: 10.1037/0022-3514.72.5.1034

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: John Wiley and sons.

Vrij, A., Fisher, R. P., and Blank, H. (2015). A cognitive approach to lie detection: a meta-analysis. *Legal Criminol. Psychol.* doi: 10.1111/lcrp.12088

Zimbler, M., and Feldman, R. S. (2011). Liar, liar, hard drive on fire: how media context affects lying behavior. *J. Appl. Soc. Psychol.* 41, 2492–2507. doi: 10.1111/j.1559-1816.2011.00827.x

Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). "Verbal and nonverbal communication of deception," in *Advances in Experimental Social Psychology*, Vol. 14, ed. L. Berkowitz (New York, NY: Academic Press), 1–57.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 June 2016
doi: 10.3389/fpsyg.2016.00814

Music As a Sacred Cue? Effects of Religious Music on Moral Behavior

Martin Lang[1,2]*, *Panagiotis Mitkidis*[3,4,5], *Radek Kundt*[2], *Aaron Nichols*[3], *Lenka Krajčíková*[6] *and Dimitris Xygalatas*[1,2,4]

[1] Department of Anthropology, University of Connecticut, Storrs, CT, USA, [2] LEVYNA Laboratory for the Experimental Research of Religion, Department for the Study of Religions, Masaryk University, Brno, Czech Republic, [3] Center for Advanced Hindsight, Social Science Research Institute, Duke University, Durham, NC, USA, [4] Interacting Minds Centre, Department of Culture and Society, Aarhus University, Aarhus, Denmark, [5] Interdisciplinary Centre for Organizational Architecture, Department of Management, Aarhus University, Aarhus, Denmark, [6] Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic

Religion can have an important influence in moral decision-making, and religious reminders may deter people from unethical behavior. Previous research indicated that religious contexts may increase prosocial behavior and reduce cheating. However, the perceptual-behavioral link between religious contexts and decision-making lacks thorough scientific understanding. This study adds to the current literature by testing the effects of purely audial religious symbols (instrumental music) on moral behavior across three different sites: Mauritius, the Czech Republic, and the USA. Participants were exposed to one of three kinds of auditory stimuli (religious, secular, or white noise), and subsequently were given a chance to dishonestly report on solved mathematical equations in order to increase their monetary reward. The results showed cross-cultural differences in the effects of religious music on moral behavior, as well as a significant interaction between condition and religiosity across all sites, suggesting that religious participants were more influenced by the auditory religious stimuli than non-religious participants. We propose that religious music can function as a subtle cue associated with moral standards via cultural socialization and ritual participation. Such associative learning can charge music with specific meanings and create sacred cues that influence normative behavior. Our findings provide preliminary support for this view, which we hope further research will investigate more closely.

Keywords: religion, music, associative learning, morality, priming

OPEN ACCESS

Edited by:
Guy Hochman,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Mark Aveyard,
American University of Sharjah, UAE
Brendan Strejcek,
Northwestern University Kellogg School of Management, USA

***Correspondence:**
Martin Lang
martin.lang@uconn.edu

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *06 November 2015*
Accepted: *17 May 2016*
Published: *07 June 2016*

Citation:
Lang M, Mitkidis P, Kundt R, Nichols A, Krajčíková L and Xygalatas D (2016) Music As a Sacred Cue? Effects of Religious Music on Moral Behavior. Front. Psychol. 7:814. doi: 10.3389/fpsyg.2016.00814

INTRODUCTION

Much psychological research conducted over the past decade has attempted to further scientific understanding of morality and ethical behavior by observing how environmental cues can enhance or degrade ethical behavior (Shariff and Norenzayan, 2007; Mead et al., 2009; Mazar and Zhong, 2010; John et al., 2014). Inferred social norms (Gino et al., 2009), ethical reminders (Mazar et al., 2008), and even decorative objects in a room (Krátký et al., 2016), have all been observed to affect dishonest behavior. This evidence suggests that automaticity plays an important role in moral decision-making based on perceptual cues (Bargh et al., 2012; Newell and Shanks, 2014). Making internalized norms salient via contextual cues can push people toward normative behavioral strategies (Cialdini et al., 1990; Hirsh et al., 2011), often without a conscious link between the two

(Bargh and Morsella, 2008). As such, behavioral responses to moral dilemmas might result from the interplay between individual norms and contextual percepts, especially in a structured environment that is regulated by normative institutions (Graham et al., 2012).

A prime example of such a normative institution is religion. Religions often strongly impact the individual's socialization process, and through the use of reminders such as symbols and repeated rituals make group-specific norms salient (Durkheim, 1912; Norenzayan and Shariff, 2008; Xygalatas, 2013). Research in recent years has shown that religious situational factors enhance the saliency of norms and play a significant role in moral decision-making (for an extensive meta-analysis see Shariff et al., 2016). However, despite an ample body of research on religious prosociality, the effects of religious contextual cues on unethical behavior are less well-documented. Only a handful of studies have looked at the effects of religious cues on deterring cheating (Bering et al., 2005; Randolph-Seng and Nielsen, 2007; Mazar et al., 2008; Piazza et al., 2011). For example, Mazar et al. (2008) found lower cheating rates amongst participants who were asked to recall the 10 Commandments compared to those who had to recall 10 book titles. Similar results were observed when using other environmental cues, such as the Islamic call to prayer (Aveyard, 2014).

These studies suggest that people modify their decisions in response to sacred cues, similarly to the way they respond to other environmental cues (for instance light in the room—Zhong et al., 2010), and that religious environments might have complex effects on people's social behavior. However, the exact mechanisms underlying the perceptual-behavioral links that affect decision-making under the influence of sacred cues are still not fully understood. Researchers have traditionally primed concepts of spirituality implicitly through the use of religiously infused anagrams (Srull and Wyer, 1979). For example, "dessert divine was fork the" would be unscrambled by participants to "the dessert was divine" (Shariff and Norenzayan, 2007). Such priming can carry semantic associations with moral norms and might also invoke fear of supernatural punishment thereby inhibiting immoral behavior. Similarly, anthropomorphic depictions of eyes might evoke a feeling of being observed and trigger reputational concerns (Bateson et al., 2006; Krátký et al., 2016). But would the same effects on moral behavior hold for arbitrary stimuli associated with religion, for instance, specific objects, gestures, or music? While the meanings of words are formed during the process of early socialization, and associations with specific actions are reinforced by everyday use, religious symbols are often confined to specific domains of one's life. Their tentative influence on moral decisions is moderated by associative learning, but it is not yet clear whether such influence would be strong enough to deter cheating. Could religious environments affect moral behavior through the accumulation of arbitrary, subtle sensory cues associated with morality?

To answer this question, we suggest a novel approach to religious priming. We selected a stimulus that does not bear any inherent meaning by itself: instrumental music. While religions employ multiple symbols that could have been chosen, music is a widespread feature of religious environments that can be translated between different cultures (as opposed to specific symbols like Shiva lingam, Christian crosses, etc.). Moreover, numerous researchers have noted that music can play a significant role in social cohesion and cooperative behavior (Kirschner and Tomasello, 2010; Dunbar et al., 2012; Pearce et al., 2015; Lang et al., 2015b). It has been suggested that music can function as a proto-symbolic system that encompasses the structure of rituals, and that religious environments might have complex effects on people's social behavior (Alcorta and Sosis, 2005). Indeed, such a connection can be described as extra-musical meaning (Koelsch, 2011) or culturally enactive meaning (Cross and Morley, 2008), referring to explicit and conventional associations of music with real-world situations (e.g., anthems making people aware of their identity; Brown, 2000). This association may work similarly to the association with linguistic concepts. In an EEG study by Koelsch et al. (2004), participants were primed with sentences or musical excerpts that were semantically either related or unrelated to a word that followed. The authors recorded an event-related brain potential that is sensitive to a semantic fit (N400) and found no difference between sentences and musical excerpts. That is, when musical excerpts were semantically unrelated to the words that followed, the same error occurred as in the case of sentences. This result suggests that music can convey linguistic concepts and prime the meaning of a word (Koelsch, 2010). Such primes have been used, for instance, in a study of purchasing behavior, showing that when music is associated with information congruent with an advertised product, participants are more likely to be persuaded by the advertisement (North et al., 2004).

Besides the extra-musical meaning, musical stimuli carry information and messages that can elicit specific emotional responses, which in turn affect mood (Thompson et al., 2001) and morality judgments (Seidel and Prinz, 2013). For example, musical stimuli with positive valence decrease concerns regarding immoral messages and increase compliance with a request to harm others (Ziv et al., 2012; Ziv, 2015). Negatively valenced music, on the other hand, can increase participants' critical thinking (Sinclair et al., 2007). Furthermore, it has been shown that the tempo of musical stimuli can influence emotional arousal (Webster and Weir, 2005) and cognitive performance (Schellenberg, 2005; Schellenberg et al., 2007). However, we lack robust evidence showing that music influences participants' actual moral behavior (Ziv et al., 2012). And if it does, does this happen via the induction of specific emotions, through an association with conceptual complexes, or both?

The current study explored whether priming participants with instrumental religious music would decrease the rate of dishonest behavior. To isolate the effects of religious music, we designed three conditions: religious, secular, and control. After exposure to one of the three stimuli, participants' task was to solve a set of 20 matrices, and for each correctly solved matrix they received a monetary reward (Mazar et al., 2008). The number of correctly solved matrices was self-reported, thereby giving participants an opportunity to report dishonestly to increase their monetary reward and inflate their performance. We predicted that participants in the religious condition would behave less dishonestly than in the other two conditions. However, because

instrumental religious music is not universally recognized as sacred (compared to religious concepts) and is thus less salient, we also expected that the effect of religious music would be moderated by participants' religiosity (congruent with the extra-musical meaning). That is, only religious participants would respond to this environmental cue that should activate an internalized behavioral schema (honesty). An additional supplementary hypothesis assumed the moderating effects of ritual participation frequency. The emotional characteristics, tempo, and impact of the presented stimuli were also assessed in order to test the hypothesis that music can affect decision-making through its affective component.

Addressing current debates on the generalizability of psychological studies (Henrich et al., 2010) and criticisms of religious priming literature and related meta-analytical research (Gomes and McCullough, 2015; van Elk et al., 2015; Shariff et al., 2016), we collected data from three different samples: a general population sample in Mauritius, and student population samples in the Czech Republic and the USA. By diversifying our participant pool, our goal was to control for possible culturally unique responses to religious primes. Despite demographic differences between these sites, we did not expect that priming with religious music should have different effects. We hypothesized that the learned link between religion and morality should work similarly in all sites. We were also interested to see whether general religiosity rates might play an important role in the effectiveness of religious primes, and we thus selected these three countries due to their different rates of general religiosity (Zuckerman, 2007; Gervais et al., under review).

MATERIALS AND METHODS

Participants

Data were collected from May 2014 to July 2015 in three sites: we recruited participants from the general Hindu population in Point aux Piments in Mauritius; a student population at Masaryk University in the Czech Republic; and a student population at Duke University, North Carolina, USA. Across the three sites, 254 participants were randomly assigned to one of three conditions: religious, secular, and control. Participants who previously took part in a similar experiment or showed suspicion about the experiment's goals were excluded from the final analysis (5 in Mauritius, 4 in the Czech Republic, and 13 in the USA). Overall, we tested 73 participants in Mauritius (20 females; $M_{age} = 30.29$, $SD = 12.95$); 78 participants in the Czech Republic (40 females; $M_{age} = 24.05$, $SD = 3.69$); and 81 in the USA (47 females; $M_{age} = 22.74$, $SD = 3.77$). Participants who did not fill out the parts of our questionnaire regarding musical stimuli ($n = 12$) were retained in the analysis of behavioral data, but were omitted from the analysis of musical stimuli. Participants were tested alone in rooms that contained only a chair, table, and computer. All materials, questionnaires, and consent forms were translated into the local languages (Mauritian Creole, Czech, and English). Informed consent was obtained from all participants. The study was approved by the Institutional Review Boards of Masaryk University, University of Connecticut, and Duke University.

Material

In a double-blind design, participants were randomly assigned to one of three conditions defined by the type of stimulus they were exposed to: religious, secular, or control. Because we were specifically interested in the effects of music, none of the used musical excerpts contained any lyrics. All stimuli were of identical duration (2 min) and were administered via headphones in order to prevent interference from external noise. In the control condition, participants were exposed to white noise in order to control for possible effects of sound manipulation. While the control stimulus was the same across the three sites (Audio 7 in Supplementary Material), music in the religious and secular conditions was site-specific.

In Mauritius, we selected the appropriate religious music after consulting local religious experts, and comparable secular music after discussing with local research assistants. For the religious condition, we chose music that is often played during collective rituals in the local temple, and in particular during the annual religious festival of Thaipusam Kavadi (Audio 1 in Supplementary Material). This musical piece has dominant fast drums and a flute sound that is characteristic of the Kavadi ritual. For the secular condition we chose a popular Bollywood song (Mera Mahi Bada Sohna Hai—"Dhaai Akshar Prem Ke"; Audio 2 in Supplementary Material) that had similar sound and tempo to the music in the religious condition by sampling the first minute without any lyrics. This minute was looped in order to create a 2 min music sample.

In the Czech Republic and the USA, we pre-screened four Christian religious songs that are used during Catholic mass and four comparable secular songs. Participants from the Czech Republic and the USA rated them on 14 characteristics. These characteristics were combined into measures of stimuli's positivity, negativity, holiness, tempo, and impact (see Supplementary Material 1.1,1.2; and Tables S1, S2). In order to select secular stimuli that would be comparable with religious stimuli, we compared the most holy stimulus with the four pre-selected secular stimuli on the ratings of positivity, negativity, tempo, and impact, and selected the least different secular stimulus.

In the Czech Republic, 40 students from Masaryk University rated the eight selected stimuli. Since all of the religious songs had similar ratings of holiness (ranging from 4.28 to 4.43 out of 6), we chose the one that had the least mean difference in all other ratings with a secular song. Using this procedure, we selected an organ version of J. S. Bach's *Ave Maria* interpreted by Charles Gounod as the religious song ($M_{holy} = 4.33$, $SD = 1.54$; Audio 3 in Supplementary Material), and Tchaikovsky's *Romance for piano in F Minor, Op 5* as the secular song (Audio 4 in Supplementary Material). *Ave Maria* was performed on organs and Tchaikovsky's piece on piano, and both songs had similar tempos. The same procedure was used in the USA to select appropriate stimuli. We used Amazon Mechanical Turk to recruit 102 participants who rated the same songs as participants in the Czech Republic. For the Religious condition, we selected J.S. Bach's *BWV 147 Jesu joy of man's desiring*, which was rated as the most sacred song ($M_{holy} = 2.94$, $SD = 2.10$; Audio 5 in Supplementary Material). The most similar secular song was

J.S. Bach's *BWV 140 Sleepers Wake* (Audio 6 in Supplementary Material). Although both songs were from the same composer, the religious one was performed on organs, while the secular one on piano.

Procedure

Our experiment was conducted using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN, and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience. Cogent 2000 was run as a Matlab Toolbox (2013a; MathWorks Inc., Massachusetts, USA). Participants were seated in individual rooms in front of a table with a computer, and a local research assistant explained that the purpose of the study was to investigate the effects of music on cognitive performance. Subsequently, the research assistant made sure that every participant understood the instructions (a practice item was presented) and instructed participants to keep their headphones on for the rest of the experiment. The research assistant then left the room, informing the participant that she or he would be working in the adjacent room and could be called when needed. The condition-specific musical stimulus played for 2 min, after which low-volume white noise was played for the rest of the experiment. This served to eliminate any possible disturbing noises.

Once the music ended, a series of mathematical tasks appeared on the screen. The participants' task was to solve as many as they could out of a total of 20 given matrices (adapted from Mazar et al., 2008). Each matrix was presented on the screen in the form of a 3 × 3 table of numbers (see **Figure 1**). In each matrix, participants had to find two numbers that added up to 10 and remember their coordinates. There was always only one correct solution. Each matrix was presented for 15 s, after which participants had 6 s to think about the correct solution. Subsequently, the correct answer appeared on the screen for 3 s, and if it matched the solution that participants had in mind, they would make a mark on a prepared sheet. The prepared sheet contained one previously filled-out row, suggesting how many matrices the previous participant had successfully solved. Because almost no cheating was observed in a pilot that was run in the Czech Republic before the current study, we decided to encourage participants to cheat by suggesting that a previous participant cheated as well (Gino et al., 2009). Thus, the pre-filled row always contained eight marks. The matrix-solving task lasted 8 min in total.

A

	1	2	3
A	5	8	7
B	1	4	0
C	1	4	2

B

	1	2	3
A	2.5391	6.2664	5.5215
B	4.5095	7.4819	3.9655
C	6.0555	4.5395	3.7336

FIGURE 1 | A comparison of easy (A) and difficult matrices (B) used in the experiment.

After participants went through all 20 matrices they were instructed by the program to call the research assistant who then administered a post-study questionnaire and compensated participants based on their self-reported number of correctly solved matrices. The questionnaire assessed participants' religiosity, familiarity with the musical piece, ratings of the stimuli's positivity, negativity, holiness, tempo, and impact, and contained basic demographics (see Supplementary Material 1.4). We used the same approach to the construction of the stimuli's measures as during pre-screening the stimuli (see Supplementary Material 1.3). Debriefing was performed at the end of data collection.

For each correctly solved equation, participants were paid 5 MUR/10 CZK/0.5 USD. The maximum possible amount that participants could earn in each site was roughly equivalent to a budget restaurant meal. We did not control how many equations participants really solved correctly. However, in contrast to Mazar et al. (2008) who used the overall number of claimed matrices in their analyses, we approached the approximation of actual cheating in a more robust way. Using the raw untransformed data would introduce variability where, in theory, there should not be any. In other words, two participants might have solved three and five matrices respectively, seemingly showing variability in cheating while actually having chosen the same behavioral strategy (honesty). Whereas this problem could be addressed with a large sample, adding predictors at the level of an individual (e.g., religiosity) could bias a predictor's explanatory power. Furthermore, using the raw data would inflate the cheating scale and any differences in cheating would appear smaller than they were in reality.

To approximate the actual levels of cheating, we designed the experiment in such a way that most participants would solve five matrices. The first two equations were easy enough that everyone who passed the comprehension test should solve (adding up two numbers from 1 to 9), while the third matrix included numbers with three decimals, making it possible to solve in 15 s. In the rest of the matrices, the numbers contained 4 or 5 decimals, making it very difficult to solve in 15 s. However, participants could also guess the correct answer with a chance of 1:36 in each of the 17 remaining equations. According to the Bernoulli probability distribution, there is a 99% probability that a participant will not guess more than two solutions correctly. We thus assumed that participants who reported five or fewer solved equations were honest (i.e., possibly solving three and guessing two matrices).

To test this assumption, we recruited 112 participants from a student population at Masaryk University in the Czech Republic who were presented with the matrix task during a lecture in a large classroom. The matrices were projected on a wall and participants were instructed to write down answers (coordinates

of two numbers adding up to 10) on a piece of paper. Participants who did not answer correctly any of the first two matrices without decimals were removed from the subsequent analysis ($n = 12$; such participants would not pass a comprehension test in our experiment), where we computed the average number of correctly solved matrices. Although the correspondence of pretest results with our assumption would not mean that we measured actual cheating, we believe that a low SD of pretest results together with a relatively large sample size provides sufficient precision for assessing the effects of our manipulation on participants' behavior.

Data Analysis

All data were analyzed in R (version 3.2.3, R Core Team, 2014). Since our data were bounded on the possible amount of dishonesty, we considered four different models: normal, normal censored, beta, and zero-inflated beta. While the untransformed data on cheating looked almost normally distributed (albeit leptokurtic; see Figure S2A), the values between 2 to 5 solved matrices mask the actual censoring. Since we considered five or fewer reported matrices as ethical behavior (see section Procedure), these values were collapsed to zero unfairly reported matrices. Thus, when this boundary was taken into account, histogram data showed a significant positive skew (see Figure S2B). Although the zero-inflation could be modeled by a censored model with normal distribution, the rest of the distribution (from the value of one and higher) was not normal either. We therefore considered a beta regression that uses a logit link to model means and variation in order to account for heteroscedasticity and skewness often present in bounded data (Stasinopoulos and Rigby, 2007; Cribari-Neto and Zeileis, 2010). To test whether a model with beta distribution would better fit the data, we transformed the number of claimed matrices to a percentage, with 15 being 100%—maximal dishonest behavior. Because our data also contained extreme values of 0 and 1 that are unacceptable for a beta regression model, we transformed the dependent variable using the formula ($y' = (y \cdot (n - 1) + 0.5)/n$), where y is the transformed variable and n is the sample size (Smithson and Verkuilen, 2006). For the beta zero-inflated model, we used percentage data without transforming 0 and 1. A difference in Akaike Information Criterion (AIC) was used to compare models with different distributions (modeling only the intercept). From the four considered models, the one with beta distribution had significantly lower AIC than the other models ($AIC_{beta} = -137.24$, $AIC_{normal} = 37.05$). Thus, we used beta regression on the transformed data to model our dependent variable.

We fitted a beta regression model (Smithson and Verkuilen, 2006; Eskelson and Madsen, 2011) using the function *gamlss* (*gamlss* package; Stasinopoulos and Rigby, 2007). We built four sets of models. In the first set, we kept site as an independent factor in all models, controlling for differences between our sites. First, we modeled the main condition effect across all sites; subsequently, we added a Condition*Religiosity interaction to the model and compared it with a model that included a Condition*Ritual participation interaction; and lastly, we added possible covariates. In the first addition, age and sex were considered. The second addition comprised of the stimuli's positivity, negativity, tempo, and impact. In the second set of models, we analyzed condition effects and a Condition*Religiosity interaction at each site. In the third set, we considered covariates that could explain tentative differences between the sites. Namely, we looked at between-site differences in religiosity; ritual participation frequency; perceived holiness of the religious stimuli; perceived negative and positive emotional valence of the stimuli; and perceived tempo and impact of the stimuli. Finally, in the fourth set, we looked at the musical characteristics of the religious stimuli and their predictive power regarding unethical behavior in the religious condition. In all models with condition effects, we set the religious condition as a reference category for comparisons. That is, we were interested only in differences between the religious condition and the other two conditions. We assumed there should be no differences between the secular and control conditions. For the models of cheating that included site as a predictor, the USA was set as the reference category, but this choice was arbitrary. Specific between-site differences in overall cheating were not of interest in the current study—we used site only as a control for effects that were outside of our interest.

RESULTS

Pretest

Results from the pretest confirmed our assumption that people on average solve five matrices ($n = 100$, $M = 4.53$, $SD = 1.57$). The minimum number of solved equations was two, while the maximum was nine. Although this range seems high at first, the frequency of participants that solved more than five matrices is exponentially decreasing (see Figure S1). We decided to set the cut-off at five as suggested by the mean number of solved matrices and Bernoulli probability distribution (see Procedure). In other words, we treated all participants in our experiment as behaving ethically if they reported five or fewer solved matrices. Six or more reported matrices were regarded as a scale of cheating.

Manipulation Check

An analysis of the perceived holiness of the stimuli across the three sites revealed a significant difference between conditions [$F_{(2, 217)} = 20.63$, $p < 0.001$]. Specifically, the religious condition had significantly higher ratings than the secular condition, and the control condition (ps < 0.001; see **Table 1** for descriptive statistics). Looking at the emotional valence of the stimuli [$F_{(2, 217)} = 4.64$, $p = 0.010$], we found that the religious condition was perceived as significantly less negative than the control condition ($p = 0.047$). We did not observe any significant differences between the religious and secular conditions ($p = 0.347$). These results were replicated also for the positivity of stimuli [$F_{(2, 217)} = 18.06$, $p < 0.001$]: the religious stimuli were rated as significantly more positive compared to the control stimuli ($p < 0.001$), but not compared to the secular stimuli ($p = 0.573$). Similar results were obtained for our measures of tempo [$F_{(2, 217)} = 6.90$, $p = 0.001$] and impact [$F_{(2, 217)} = 4.97$, $p = 0.008$] of the stimuli. The religious stimuli were rated as significantly slower than the control stimuli ($p = 0.001$),

but there was no difference between the religious and secular stimuli ($p = 0.874$). In terms of impact, the religious condition had significantly higher impact than the control condition ($p = 0.002$). The difference between the religious and secular condition was not significant ($p = 0.219$).

Dishonest Behavior

To assess the amount of dishonest behavior among participants, we measured the percentage of matrices that were claimed as correctly solved and used beta regressions to estimate differences between predictors. We did not observe a significant difference between the religious and the secular ($p = 0.44$) and control conditions ($p = 0.14$). The estimates with significance levels from a beta regression are displayed in **Table 2**, Model 1 and plotted in **Figure 2A**. Looking at differences between the sites, participants in Mauritius claimed significantly more solved matrices than participants in the USA ($p = 0.007$), while participants in the Czech Republic claimed significantly fewer ($p = 0.004$; **Table 2**, Model 1). We observed a significant Condition*Religiosity interaction, with religious people cheating significantly less in the religious condition ($p = 0.027$). Compared to the religious condition, religiosity played a significantly smaller role in the secular ($p = 0.026$) and control conditions ($p = 0.039$; see **Table 2**, Model 2 and **Figure 2B**). That is, the more religious participants were, the less they cheated in the religious condition, while in the other two conditions religiosity did not significantly affect cheating. The model comprising a Condition*Ritual participation interaction suggested the same

TABLE 1 | Descriptive statistics of dishonest behavior and musical-stimuli ratings.

Variable	Religious ($n = 74$)				Secular ($n = 80$)				Control ($n = 78$)			
	M	SD	CI	*d*	M	SD	CI	*d*	M	SD	CI	*d*
% Claimed	30.27	27.35	24.04–36.05	–	31.50	24.41	26.37–36.63	0.05	34.96	27.71	28.81–41.12	0.17
Holiness	3.84	1.58	3.47–4.21	–	2.86	1.32	2.56–3.15	0.68	2.42	1.16	2.15–2.68	1.03
Negativity	2.28	0.92	2.10–2.49	–	2.13	0.80	1.95–2.31	0.17	2.59	1.09	2.34–2.84	0.31
Positivity	3.11	0.84	2.91–3.31	–	3.20	0.89	2.99–3.40	0.10	2.34	1.10	2.09–2.59	0.78
Tempo	2.73	0.96	2.50–2.95	–	2.76	0.83	2.57–2.94	0.03	3.23	0.96	3.01–3.45	0.52
Impact	3.26	1.13	3.01–3.53	–	3.01	1.28	2.73–3.30	0.21	2.63	1.63	2.34–2.91	0.53

CI = 95% Confidence intervals. Cohen's d is the effect size of comparisons between the religious condition and the other conditions.

TABLE 2 | Estimates with SE from beta regressions for the percentage of matrices claimed as correct.

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	29.84 (3.61)***	30.32 (6.27)***	30.27 (6.19)***	31.18 (6.78)***	29.86 (6.49)***
Mauritius	11.32 (4.18)**	11.10 (4.44)*	9.28 (4.21)*	7.93 (4.77)†	9.27 (5.19)†
Czech Republic	−9.61 (3.34)**	−9.63 (3.44)**	−9.38 (3.39)*	−10.50 (3.48)**	−9.75 (4.20)*
Secular	3.04 (3.93)	2.45 (3.92)	3.48 (3.95)	2.74 (3.97)	3.03 (3.93)
Control	5.99 (4.07)	5.40 (4.05)	6.01 (4.06)	6.19 (4.16)	7.60 (4.38)†
Religiosity		−5.31 (2.40)*		−4.97 (2.48)*	−4.97 (2.43)*
Secular*Religiosity		7.55 (3.37)*		7.54 (3.45)*	7.32 (3.37)*
Control*Religiosity		6.60 (3.18)*		6.49 (3.28)*	6.26 (3.18)*
Ritual			−1.55 (1.47)		
Secular* Ritual			5.40 (2.12)*		
Control* Ritual			3.54 (2.06)†		
Females vs. Males				7.90 (3.50)*	8.47 (3.48)*
Age				0.04 (0.20)	0.04 (0.20)
Positivity					−2.16 (2.24)
Negativity					−2.70 (2.12)
Tempo					−1.61 (1.84)
Impact					1.99 (1.86)
Cox-Snell R^2	0.124	0.147	0.157	0.166	0.175

*In all models, we control for the effects of site. The religious condition and the USA site were set as reference categories (intercept). The first model contains only the effects of condition (compared to the religious condition) while controlling for the effects of site. The second model includes a Condition*Religiosity interaction, describing the effects of religiosity on cheating in the religious condition. The two predictors specified as interactions (Secular*Religiosity and Control*Religiosity) are comparisons with this effect. Again, we control for site. The third model has an identical design to the second, only with a Condition*Ritual participation interaction. Since the effects of ritual participation on morality were not as strong as those of religiosity, we retained the latter factor for subsequent models. The fourth model contains site and condition effects, the significant interaction, and demographic covariates. The fifth model controls also for different characteristics of our musical stimuli.*

†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

trend (for religious condition, $p = 0.294$), but the interaction was significant only for the secular condition ($p = 0.011$) and not for the control condition ($p = 0.086$; see **Table 2**, Model 3). From the considered covariates, only sex significantly improved the model fit. Aggregating across the three sites, on average males reported more matrices than females ($p = 0.025$; see **Table 2**, Model 4). There was no effect of perceived valence ($p_{negativity} = 0.203$; $p_{positivity} = 0.335$; $p_{tempo} = 0.382$; $p_{impact} = 0.286$) of the stimuli or of age ($p = 0.847$) on participants' behavior (**Table 2**, Model 5).

Between-Sites Differences

Focusing on the differences between our three sites (Mauritius, the Czech Republic, and the USA), we built separate models for the condition effects (see **Table 3** and **Figure 3** for descriptive statistics and **Table 4** for model estimates). First, there was a significant difference between the religious condition and the other two conditions in Mauritius. Specifically, participants in the religious condition claimed a lower percentage of solved matrices than participants in the secular condition ($p = 0.043$) and participants in the control condition ($p = 0.044$). We did not observe a significant main effect of condition in the Czech Republic (religious vs. secular: $p = 0.581$; religious vs. control: $p = 0.891$). Likewise, the condition effect was not significant in the USA (religious vs. secular: $p = 0.718$; religious vs. control: $p = 0.695$). Looking at the Condition*Religiosity interactions, we observed a marginally significant interaction in the USA sample (Religiosity*Secular: $p = 0.068$; Religiosity*Control: $p = 0.052$), but this interaction did not replicate in the other sites (ps > 0.3; **Table 4**, Models B).

In order to better understand why the results from Mauritius differed from the other two sites, we used site as an independent variable (with Mauritius as the reference category) in predicting religiosity and ritual participation; and holiness, tempo, impact, and valence of the religious stimuli (see **Table 5** for descriptive statistics). Mauritian participants reported being significantly more religious [$F_{(2, 229)} = 13.31$, $p < 0.001$] than those in the Czech Republic ($p = 0.003$) and the USA ($p < 0.001$). Similarly, participants in Mauritius reported significantly more frequent ritual participation [$F_{(2, 229)} = 14.41$, $p < 0.001$] compared to participants in the Czech Republic ($p < 0.001$) and the USA ($p = 0.010$). Religiosity and ritual participation are plotted in **Figure 4**.

There were no significant differences [$F_{(2, 67)} = 1.03$, $p = 0.364$] between Mauritius and the other sites in perceived holiness of the religious stimuli (Czech Rep.: $p = 0.370$; USA: $p = 0.157$). However, there were significant differences in perceived negativity of the religious stimuli [$F_{(2, 67)} = 20.55$, $p < 0.001$], with the Mauritian stimulus rated as significantly more negative compared to the USA ($p < 0.001$), but not to the Czech Republic ($p = 0.592$). Conversely, this pattern of significance was reversed for the positivity of the religious

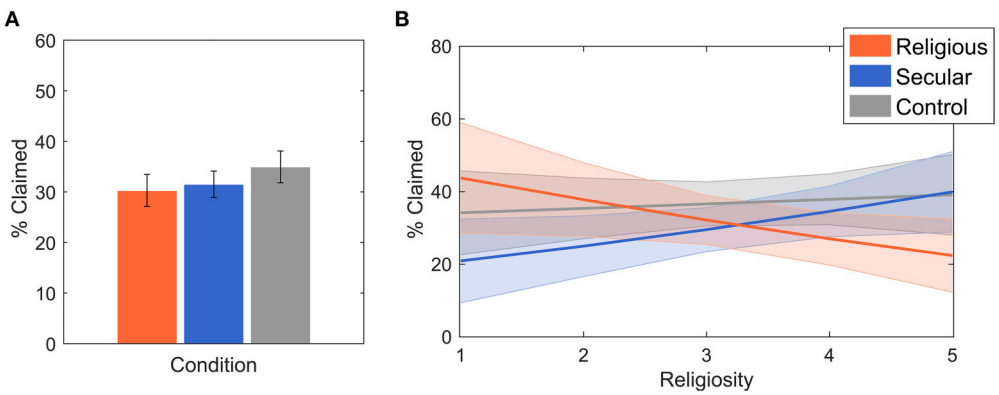

FIGURE 2 | (A) The effects of different stimuli on the percent of matrices that were claimed as correctly solved above the expected levels with ±SEM. While controlling for the effects of site, there were no significant differences between conditions (see **Table 2**, Model 1). **(B)** Predicted values with 95% confidence intervals for the Condition*Religiosity interaction. The significantly different slopes suggest that religious participants cheated less upon being exposed to religious music (**Table 2**, Model 2).

TABLE 3 | Descriptive statistics of between-sites differences in dishonest behavior (% Claimed).

	Religious					Secular					Control				
Site	***n***	**M**	**SD**	**CI**	***d***	***n***	**M**	**SD**	**CI**	***d***	***n***	**M**	**SD**	**CI**	***d***
Mauritius	21	36.83	32.91	22.75–50.90	–	25	46.67	22.36	37.90–55.43	0.35	27	49.83	30.11	38.02–60.74	0.40
Czech Rep.	27	21.73	19.27	14.46–28.30	–	27	20.00	22.57	11.49–28.51	0.08	24	20.56	18.43	13.18–27.93	0.06
USA	26	33.85	28.34	22.95–44.74	–	28	29.05	17.80	22.45–35.64	0.20	27	33.33	25.62	23.67–42.00	0.02

CI = 95% Confidence intervals. Cohen's d is the effect size of comparisons between the religious condition and the other conditions.

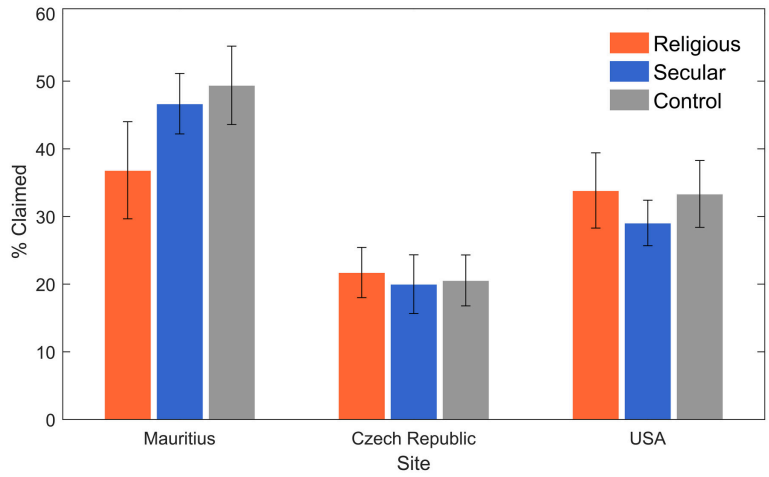

FIGURE 3 | The condition effect divided by site with ±SEM. The only significant differences between conditions were found in Mauritius.

TABLE 4 | Estimates with SE from beta regressions for the percentage of matrices claimed as correct across our three sites.

	Mauritius		Czech Republic		USA	
	Model A	**Model B**	**Model A**	**Model B**	**Model A**	**Model B**
Intercept	33.89 (5.57)**	35.49 (15.24)**	20.20 (3.57)***	21.70 (3.50)***	33.82 (5.09)**	32.46 (9.78)**
Secular	16.66 (8.10)*	17.13 (11.33)	−2.54 (4.43)	−4.29 (4.46)	−3.17 (6.74)	0.49 (6.77)
Control	16.34 (7.95)*	13.83 (8.27)†	0.93 (4.88)	0.23 (4.83)	1.01 (7.03)	4.48 (6.97)
Religiosity		−4.72 (5.23)		−4.40 (3.22)		−5.04 (3.60)
Secular*Religiosity		2.41 (9.87)		1.58 (4.20)		9.91 (5.36)†
Control*Religiosity		7.58 (7.52)		−0.57 (3.81)		9.89 (5.01)†
Cox-Snell R^2	0.071	0.085	0.008	0.077	0.005	0.068

*Models A describe condition effects for the three sites: Mauritius, the Czech Republic, and the USA. Models B display a Condition*Religiosity interaction for each site. In all models, the religious condition was set as a reference category.*

*†p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 5 | Descriptive statistics of between-sites differences in religiosity and religious-stimuli ratings.

	Mauritius (n = 73)				Czech Republic (n = 78)				USA (n = 81)			
Variable	**M**	**SD**	**CI**	***d***	**M**	**SD**	**CI**	***d***	**M**	**SD**	**CI**	***d***
Religiosity	3.81	0.89	3.60–4.01	–	3.30	1.09	3.05–3.54	0.51	2.89	1.28	2.61–3.17	0.84
Ritual participation	4.21	1.59	3.84–4.57	–	2.65	1.73	2.27–3.04	0.94	3.33	1.98	2.90–3.76	0.49
Holiness	3.41	2.00	2.46–4.36	–	3.85	1.32	3.35–4.35	0.26	4.12	1.51	3.54–4.69	0.40
Negativity	2.78	0.76	2.42–3.14	–	2.66	0.89	2.32–2.99	0.15	1.55	0.50	1.36–1.74	1.93
Positivity	2.59	0.69	2.26–2.92	–	3.55	0.73	3.27–3.83	1.35	2.99	0.83	2.68–3.31	0.53
Tempo	3.15	1.21	2.57–3.72	–	2.30	0.72	2.02–2.57	0.85	2.90	0.85	2.58–3.23	0.23
Impact	2.88	1.10	2.36–3.40	–	4.00	1.07	3.50–4.41	1.03	2.75	0.75	2.46–3.04	0.14

CI = 95% Confidence intervals. Cohen's d is the effect size of comparisons between Mauritius and the other sites.

stimuli [$F_{(2, 67)} = 8.83$, $p < 0.001$], with the Mauritian stimulus being significantly less positive than the stimulus in the Czech Republic ($p < 0.001$) but not compared to the stimulus used in the USA ($p = 0.093$). Similar results were obtained for the tempo [$F_{(2, 67)} = 5.37$, $p = 0.007$] and impact [$F_{(2, 67)} = 12.67$, $p < 0.001$] of the religious stimuli. The Mauritian stimulus was rated as significantly faster than the stimulus used in the Czech Republic ($p = 0.003$), but there was no significant difference between Mauritius and the USA ($p = 0.393$). The Czech religious stimulus had a

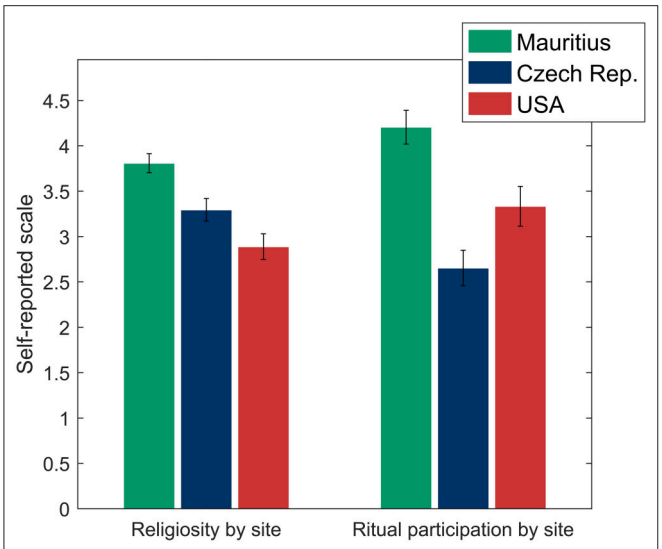

FIGURE 4 | Differences between sites in religiosity and ritual participation frequency with ±SEM. Mauritian participants were significantly more religious and attended rituals more frequently than participants in the Czech Republic and the USA.

TABLE 6 | Estimates with SE from a beta regression for the percentage of matrices claimed as correct in the religious condition.

Intercept	30.87 (5.89)***
Positivity	−0.54 (4.47)
Negativity	−0.70 (3.79)
Tempo	−3.23 (3.31)
Impact	−3.57 (3.37)
Cox-Snell R^2	0.028

Differences between sites in the characteristics of religious stimuli do not explain differences in the number of claimed matrices.

higher impact on participants compared to the Mauritian one ($p < 0.001$), but again, no significant difference was found between Mauritius and the USA ($p = 0.664$). In order to investigate whether these differences affected decision-making in the Religious condition, we built a model with the number of matrices claimed as a dependent variable, and the religious stimuli's characteristics as predictors. However, none of these characteristics explained any significant amount of variation in dishonest behavior in the religious condition (all ps > 0.29; see **Table 6**).

DISCUSSION

We tested the hypothesis that non-verbal religious primes in the form of religious music would decrease dishonest behavior compared to secular music and white noise. Whereas it has been previously shown that religious words and complex religious contexts (e.g., a church environment) can increase participants' prosociality (Xygalatas, 2013), a possible effect of religion on deterring antisocial behavior was tested only by priming with religious words. We were interested in whether moral decision-making would be influenced by such a subtle cue as instrumental music. Participants in Mauritius, the Czech Republic, and the USA were given an opportunity to dishonestly inflate their performance in order to maximize their profit. This incentive to behave dishonestly was shown to be effective across all three sites. When collapsing all three sites together, we did not observe a significant effect of religious music on the rate of dishonest behavior. However, breaking down the condition effect by site revealed that religious music significantly decreased the incentive to cheat in Mauritius, but no such effect was observed in the other two sites. To test the hypothesis that the condition effect would be moderated by religiosity, we included a Condition*Religiosity interaction in our models. Religious music significantly reduced dishonest behavior in religious participants, while ritual participation frequency played a marginally significant role in the religious condition. Males displayed higher rates of dishonesty across the three conditions. Finally, participants' age and musical characteristics of the stimuli did not play a significant role. Together, these results offer a more nuanced interpretation of the influence of religious contexts on moral behavior.

It is important to acknowledge that the current study has several limitations. First, given the effect sizes for the differences between conditions at each site, we need to exert caution in interpreting the observed differences. While the collapsed sample across all sites is robust enough to detect medium effect sizes, the sample sizes at each site do not warrant generalizations due to low statistical power (Button et al., 2013). Furthermore, since the effect sizes of the differences between conditions in Mauritius are rather small (0.3 and 0.4), this finding needs to be further probed by future studies. Second, we did not collect exact data on actual cheating. While our procedure should secure confident estimates of unethical behavior, it is still possible that some participants correctly solved more than 5 matrices and vice versa. Similarly, some participants could feel that they found a correct answer and that the answer we provided was incorrect. Since the mathematical equations were computed under time-pressure, participants could make a small mistake without noticing and feel righteous to claim their answer as correct. However, given our overall sample size, such participants should constitute only a minimal portion of our sample. Third, since the musical stimuli were played before the mathematical task, their effects could be concealed by the time delay or the cognitive demands of the task. Perhaps if the stimuli were played during the whole experiment, the primes would be more salient and thus capable of influencing participants' behavior to a greater extent. Such a proposition needs further empirical testing. Fourth, the religiosity effect could have been mediated by some other mental process than by an association to normative behavior. For example, the thought of religion could have primed global processing, which has been previously shown to increase prosocial behavior (Mukherjee et al., 2014).

The lack of a main condition effect in the overall sample suggests that religious music might not always be salient enough to deter people from dishonest behavior. Although our religious

stimuli were recognized as significantly more holy than the other two stimuli, honesty was only affected in one of three sites. A significantly lower amount of dishonest behavior in the religious condition was observed only in Mauritius, which points to the need for a more thorough understanding of differences between our sites. There are at least three possible interpretations: (a) this finding is a false positive; (b) participants in Mauritius were induced with different emotions that influenced their behavior; or (c) the association between religious music and normative behavior is stronger in Mauritius due to higher religiosity.

The observed difference between different conditions in Mauritius could have been caused by different characteristics of our religious stimuli. While we used organ music in the Czech Republic and the USA, the Mauritian religious stimulus had significantly higher tempo and dominant drums. A comparison of religious stimuli across sites revealed mixed results. The Mauritian religious music was perceived as significantly more negative than the religious stimulus in the USA, while there was no difference between Mauritius and the Czech Republic. We can speculate that, for example, Mauritian participants were more avoidant and critical due to higher negativity evoked by the religious stimulus and, consequently, avoided the cheating behavior. However, we find this interpretation unlikely because the perceived negativity of the stimuli was not significantly different between Mauritius and the Czech Republic. Similarly, differences between Mauritius and the other sites in positivity, tempo, and impact were always only between two sites, suggesting that no systematic differences were related to those properties. Furthermore, looking at the overall effects of musical characteristics on cheating rates, we did not observe any significant influence of these variables. This is in contrast with previous research which suggested that positively valenced music decreases moral concerns (Ziv et al., 2012). The lack of such effects might stem from the fact that the link between positive music and cheating was previously tested only by self-reports (Ziv et al., 2012). Alternatively, the cognitive demands of our task might have concealed any tentative subtle effects of musical characteristics.

The overall higher rates of self-reported religiosity and ritual participation frequency in Mauritius appear to be a more probable explanation of the behavioral differences between our sites. Religiosity is entrenched into Mauritian everyday life much more than in the other two sites, and might play a more important normative role (Xygalatas, 2013). This was confirmed by the significant differences in reported religiosity and frequency of ritual participation between Mauritius and the other two sites, and might indicate that higher religiosity could be associated with heightened sensitivity to religious cues (for similar results on prosocial behavior see Xygalatas et al., 2015). This interpretation is further supported by the significant Condition*Religiosity interaction. Collapsing all three sites, higher religiosity was associated with decreased rates of dishonest behavior in the religious condition. Although participants recognized our stimuli as religious, the less religious participants seemed to be unaffected. This result is in contrast with previous studies that showed no effect of religiosity on overall cheating rates (Randolph-Seng and Nielsen, 2007; Mazar et al., 2008; Aveyard, 2014). Our study thus offers new preliminary evidence on the role of religiosity, in congruence with the research on religious prosocial behavior (Shariff et al., 2016).

The fact that religiosity had a significant impact on dishonest behavior only in the religious condition supports the important role of religious situational factors in decision-making. We propose that dispositional religiosity does not affect participants' honesty to a large extent, unless it is activated by environmental sacred cues (Darley and Batson, 1973; Norenzayan and Shariff, 2008; Xygalatas, 2013; Xygalatas et al., 2015). While Mauritian participants reported significantly higher religiosity than participants at the other sites, the Mauritian cheating rates were significantly higher than those in the Czech Republic and the USA. Such a finding suggests that participants needed to be reminded of their religiosity in order for it to affect their moral decision-making. However, such a "reminder effect" is probably temporary (Malhotra, 2008) and confined only to religious participants. When religious cues are salient and general enough (e.g., the word God), they might affect non-religious participants, thus masking the effect of dispositional religiosity. But when subtle (as in the case of our study), these sacred cues only influence religious people who are more sensitive to them. This could also explain why studies that used linguistic primes (Randolph-Seng and Nielsen, 2007; Mazar et al., 2008) did not find a significant moderating effect of religiosity. Religious words are part of the standard cultural language toolbox and have stronger behavioral associations than specific religious symbols. For example, the Islamic call to prayer is a public, omnipresent cue that is directly associated with specific behaviors. As such, these cues are less ambiguous than music (Cross and Morley, 2008). Instrumental religious music, on the other hand, is generally less known, and associative learning is rather accomplished via communal socialization that reinforces the association of symbols with religion. Music is rarely associated with specific behavioral requirements, especially those regarding moral conduct. Behavioral schemas are thus not directly accessible to those who have not undergone religious socialization and do not participate in communal ritual gatherings (while they might be accessible to the majority of people through words). The fact that music is such a subtle cue can explain why we did not observe a significant Condition*Religiosity interaction in each of our sites. We would probably need larger sample sizes in order to show such an interactive effect.

The importance of ritual participation in the accessibility of behavioral schemas is further supported by a trend in the Condition*Ritual participation interaction. The fact that this trend did not reach statistical significance, however, suggests that ritual participation alone might not be enough to promote honest behavior (Mitkidis et al., 2014). It may reinforce the link between symbolic and behavioral schemas, but this link without an overarching religious worldview is probably a weak motivational force. Although participation in public rituals usually signals acceptance of religious norms (Rappaport, 1999), it is not necessarily tied to actual normative behavior and

people can participate in these rituals for various reasons, for instance, reducing anxiety (Lang et al., 2015a), including no specific reason at all (Xygalatas, 2012). Such participants might be less inclined to follow normative schemas prescribed by their respective religions, especially if different behaviors have momentarily higher pay-offs (free-riding). Furthermore, ritual intensity may play an important role in the reinforcement of the link between symbol and behavior. High-intensity rituals are usually extremely arousing events (Xygalatas et al., 2013a,b), and as such might yield stronger affective bonds between symbols and conceptual complexes (Alcorta and Sosis, 2005). This might provide additional support for the suggested explanation of the differences in dishonest behavior between our sites. In Mauritius, we used music from the Kavadi ritual as the religious stimulus. The Kavadi is a high-intensity ritual that involves multiple body piercings, walking on nails, carrying heavy objects, and other forms of prolonged suffering. As such, it might be especially powerful in associating the musical stimulus with specific behavioral requirements and might have provided sufficient motivation for moral behavior that was not reached by religious stimuli that referred to less intense rituals in the other sites. This interpretation gains additional support by field experimental evidence that self-reported frequency of participation in the Kavadi ritual significantly predicted lower amounts of dishonest behavior in an economic game (Xygalatas et al., under review). We thus suggest that participation in high-intensity rituals might be effective in transforming behavioral requirements into symbols and as such be a powerful motivational force.

Our findings might be of importance for evolutionary models of music and its functions. Evolutionary theorists have disagreed on whether music is an evolutionary by-product or an adaptation. The by-product thesis argues that music parasitizes upon our evolved language abilities. In fact, Steven Pinker (1998) has dubbed music an "auditory cheesecake." According to this view, our love for music is a by-product of specific cognitive-linguistic capacities, just like our love for junk food is a by-product of our adaptive need for fat, salt, and sugar. Others, however, point to the ubiquity of music across all cultures, as well as the fact that language and musical abilities are not strictly cognitively overlapping, and argue that music-making might have evolved as an adaptive trait (Fitch, 2006). For example, it might be an important tool for sexual selection, much like in birds (Miller, 2000), as suggested by the sex appeal of musical celebrities. Another important function might be related to an endorphin-based social binding mechanism (Dunbar et al., 2012) whereby music can function as social glue, a sort of "vocal grooming" (Weinstein et al., 2015). While these functions are not mutually exclusive, here we demonstrate that music may serve yet another function, that of representing norms and influencing behavioral schemas. We suggest that it does so via associative learning in communal gatherings where conceptual complexes are encoded in memory together with music. This link might be even stronger when norm-related words are included to create a song. Such songs can trigger outbursts of connotations, and thus function as a compact version of normative conceptual complexes, becoming effective vehicles for the transmission of social norms.

In summary, the current study provides preliminary support for the hypothesis that instrumental music can serve as a reminder of normative behavior, but only for participants who previously formed an association between religion and specific music. This result suggests that while socialization into group norms is crucial for ethical behavior, people need to be reminded of these norms to ensure an activation of normative behavioral schemes. In this respect, religion is a powerful institution that fosters normative behavior via shared rituals, repetitive songs and prayers, and other symbols that can act as associative triggers of ethical behavior. Further research should also investigate whether a combination of these triggers might possibly amplify their effects on participants' decision making. Likewise, using multiple sites within different cultural contexts in future research might help increase the reliability of priming studies and address the reproducibility crisis in psychological research (Open Science Collaboration, 2015).

AUTHOR CONTRIBUTIONS

ML, PM, RK, and DX designed the study; ML, RK, AN, and LK collected the data; ML analyzed the data; ML, PM, AN, and DX wrote the paper.

ACKNOWLEDGMENTS

We thank Guy Hochman and the two reviewers for valuable comments on previous versions of this study. We are grateful to Mijal Bucay, Will Fedder, Laura Hamon-Meunier, Anestis Karasaridis, Danijela Kurečková, and Mehreen May for help with data collection. This research was supported by the project "LEVYNA Laboratory for the Experimental Research of Religion" (CZ.1.07/2.3.00/20.048), co-financed by the European Social Fund and the state budget of the Czech Republic; the Faculty of Arts, Masaryk University; the "Technologies of the Mind" project at the Interactive Minds Centre at Aarhus University, financed by the Velux Foundation; and the Cultural Evolution of Religion Research Consortium, financed by the Canadian Social Sciences and Humanities Research Council (SSHRC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00814

REFERENCES

Alcorta, C. S., and Sosis, R. (2005). Ritual, emotion, and sacred symbols: the evolution of religion as an adaptive complex. *Hum. Nat.* 16, 323–359. doi: 10.1007/s12110-005-1014-3

Aveyard, M. E. (2014). A call to honesty: extending religious priming of moral behavior to Middle Eastern Muslims. *PLoS ONE* 9: e99447. doi: 10.1371/journal.pone.0099447

Bargh, J. A., and Morsella, E. (2008). The unconscious mind. *Perspect. Psychol. Sci.* 3, 73–79. doi: 10.1111/j.1745-6916.2008.00064.x

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., and Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends Cogn. Sci.* 16, 593–605. doi: 10.1016/j.tics.2012.10.002

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* 2, 412–414. doi: 10.1098/rsbl.2006.0509

Bering, J. M., McLeod, K., and Shackelford, T. K. (2005). Reasoning about dead agents reveals possible adaptive trends. *Hum. Nat.* 16, 360–381. doi: 10.1007/s12110-005-1015-2

Brown, S. (2000). "Evolutionary models of music: from sexual selection to group selection," in *Perspectives in Ethology, Volume 13: Evolution, Culture, and Behavior*, eds Tonneau and Thompson (New York, NY: Kluwer Academic/Plenum Publishers), 231–281.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015

Cribari-Neto, F., and Zeileis, A. (2010). Beta regression in R. *J. Stat. Softw.* 34, 1–24. doi: 10.18637/jss.v034.i02

Cross, I., and Morley, I. (2008). "The evolution of music : theories, definitions and the nature of the evidence," in *Communicative Musicality*, eds I. Cross and I. Morley (Oxford: Oxford University Press), 61–82.

Darley, J. M., and Batson, C. D. (1973). "From Jerusalem to Jericho": a study of situational and dispositional variables in helping behavior. *J. Pers. Soc. Psychol.* 27, 100–108. doi: 10.1037/h0034449

Dunbar, R. I. M., Kaskatis, K., MacDonald, I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702. doi: 10.1177/147470491201000403

Durkheim, E. (1912). *The Elementary Forms of the Religious Life*. New York, NY; London; Toronto, ON; Sydney, NSW; Tokyo; Singapore: The Free Press.

Eskelson, B., and Madsen, L. (2011). Estimating Riparian understory vegetation cover with Beta regression and copula models. *For. Sci.* 57, 212–221.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gomes, C. M., and McCullough, M. E. (2015). The effects of implicit religious primes on dictator game allocations: a preregistered replication experiment. *J. Exp. Psychol. Gen.* 144, e94–e104. doi: 10.1037/xge0000027

Graham, J., Meindl, P., and Beall, E. (2012). Integrating the streams of morality research: the case of political ideology. *Curr. Dir. Psychol. Sci.* 21, 373–377. doi: 10.1177/0963721412456842

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83; discussion 83–135. doi: 10.1017/S0140525X0999152X

Hirsh, J., Galinsky, A., and Zhong, C. (2011). Drunk, powerful, and in the dark: how general processes of disinhibition produce both prosocial and antisocial behavior. *Perspect. Psychol. Sci.* 6, 415–427. doi: 10.1177/1745691611416992

John, L. K., Loewenstein, G., and Rick, S. I. (2014). Cheating more for less: upward social comparisons motivate the poorly compensated to cheat. *Organ. Behav. Hum. Decis. Process.* 123, 101–109. doi: 10.1016/j.obhdp.2013.08.002

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* 14, 131–137. doi: 10.1016/j.tics.2010.01.002

Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Front. Psychol.* 2:110. doi: 10.3389/fpsyg.2011.00110

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197

Krátký, J., McGraw, J., Xygalatas, D., Mitkidis, P., and Reddish, P. (2016). It depends who is watching you: 3-dimensional agent representations increase generosity in a naturalistic setting, *PLoS ONE* 11:*e*0148845. doi: 10.1371/journal.pone.0148845

Lang, M., Krátký, J., Shaver, J. H., Jerotijević, D., and Xygalatas, D. (2015a). Effects of Anxiety on Spontaneous Ritualized Behavior. *Curr. Biol.* 25, 1892–1897. doi: 10.1016/j.cub.2015.05.049

Lang, M., Shaw, D. J., Reddish, P., Wallot, S., Mitkidis, P., and Xygalatas, D. (2015b). Lost in the rhythm: effects of rhythm on subsequent interpersonal coordination. *Cogn. Sci.* doi: 10.1111/cogs.12302. [Epub ahead of print].

Malhotra, D. (2008). (When) are religious people nicer? Religious salience and the'sunday effect'on pro-social behavior. *Judgement Decis. Mak.* 5, 138–143.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mazar, N., and Zhong, C.-B. (2010). Do green products make us better people? *Psychol. Sci.* 21, 494–498. doi: 10.1177/0956797610363538

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., and Ariely, D. (2009). Too tired to tell the truth: self-control resource depletion and dishonesty. *J. Exp. Soc. Psychol.* 45, 594–597. doi: 10.1016/j.jesp.2009.02.004

Miller, G. (2000). "Evolution of human music through sexual selection," in *The Origins of Music*, eds N. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 329–360.

Mitkidis, P., Lienard, P., Nielbo, K. L., and Sørensen, J. (2014). Does goal-demotion enhance cooperation? *J. Cogn. Cult.* 14, 263–272. doi: 10.1163/15685373-12342124

Mukherjee, S., Srinivasan, N., and Manjaly, J. A. (2014). Global processing fosters donations toward charity appeals framed in an approach orientation. *Cogn. Process.* 15, 391–396. doi: 10.1007/s10339-014-0602-8

Newell, B. R., and Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behav. Brain Sci.* 37, 1–19. doi: 10.1017/S0140525X12003214

Norenzayan, A., and Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science* 322, 58–62. doi: 10.1126/science.1158757

North, A., Mackenzie, L., Law, R., and Hargreaves, D. (2004). The effects of musical and voice "Fit" on responses to advertisements. *J. Appl. Soc. Psychol.* 34, 1675–1708. doi: 10.1111/j.1559-1816.2004.tb02793.x

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

Pearce, E., Launay, J., and Dunbar, R. I. M. (2015). The ice-breaker effect: singing mediates fast social bonding. *R. Soc. Open Sci.* 2:150221. doi: 10.1098/rsos.150221

Piazza, J., Bering, J. M., and Ingram, G. (2011). " Princess Alice is watching you": Children's belief in an invisible person inhibits cheating. *J. Exp. Child Psychol.* 109, 311–320. doi: 10.1016/j.jecp.2011.02.003

Pinker, S. (1998). *How the Mind Works*. New York, NY: Penguin Books.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Randolph-Seng, B., and Nielsen, M. E. (2007). Honesty : one effect of primed religious representations. *Int. J. Psychol. Relig.* 17, 303–315. doi: 10.1080/10508610701572812

Rappaport, R. (1999). *Ritual and Religion in the Making of Humanity*. Cambridge: Cambridge University Press.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Curr. Dir. Psychol. Sci.* 14, 317–320. doi: 10.1111/j.0963-7214.2005.00389.x

Schellenberg, E. G., Nakata, T., Hunter, P. G., and Tamoto, S. (2007). Exposure to music and cognitive performance: tests of children and adults. *Psychol. Music* 35, 5–19. doi: 10.1177/0305735607068885

Seidel, A., and Prinz, J. (2013). Sound morality: irritating and icky noises amplify judgments in divergent moral domains. *Cognition* 127, 1–5. doi: 10.1016/j.cognition.2012.11.004

Shariff, A. F., and Norenzayan, A. (2007). God is watching you: priming god concepts increases prosocial behavior in an anonymous economic game. *Psychol. Sci.* 18, 803–809. doi: 10.1111/j.1467-9280.2007.01983.x

Shariff, A. F., Willard, A. K., Andersen, T., and Norenzayan, A. (2016). Religious priming: a meta-analysis with a focus on prosociality. *Personal. Soc. Psychol. Rev*. 20, 27–48. doi: 10.1177/1088868314568811

Sinclair, R., Lovsin, T., and Moore, S. (2007). Mood state, issue involvement, and argument strength on responses to persuasive appeals. *Psychol. Rep.* 101, 739–753. doi: 10.2466/pr0.101.7.739-753

Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* 11, 54–71. doi: 10.1037/1082-989X.11.1.54

Srull, T. K., and Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: some determinants and implications. *J. Pers. Soc. Psychol.* 37, 1660–1672. doi: 10.1037/0022-3514.37.10.1660

Stasinopoulos, D., and Rigby, R. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *J. Stat. Softw.* 23, 1–46. doi: 10.18637/jss.v023.i07

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal, mood, and the mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345

van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., and Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Front. Psychol.* 6:1365. doi: 10.3389/fpsyg.2015.01365

Webster, G. D., and Weir, C. G. (2005). Emotional responses to music: interactive effects of mode, texture, and tempo. *Motiv. Emot.* 29, 19–39. doi: 10.1007/s11031-005-4414-0

Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I. M., and Stewart, L. (2015). Singing and social bonding: Changes in connectivity and pain threshold as a function of group size. *Evol. Hum. Behav.* 37, 152–158. doi: 10.1016/j.evolhumbehav.2015.10.002

Xygalatas, D. (2013). Effects of religious setting on cooperative behavior: a case study from Mauritius. *Religion Brain Behav.* 3, 91–102. doi: 10.1080/2153599X.2012.724547

Xygalatas, D. (2012). *The Burning Saints : Cognition and Culture in the Fire-walking Rituals of the Anastenaria*. London: Equinox.

Xygalatas, D., Kundtová-Klocová, E., Cigán, J., Kundt, R., Maňo, P., Kotherová, S., et al. (2015). Location, location, location: effects of cross- religious primes on prosocial behaviour. *Int. J. Psychol. Relig.* doi: 10.1080/10508619.2015.1097287. [Epub ahead of print].

Xygalatas, D., Mitkidis, P., Fischer, R., Reddish, P., Skewes, J., Geertz, A. W., et al. (2013a). Extreme rituals promote prosociality. *Psychol. Sci.* 24, 1602–1605. doi: 10.1177/0956797612472910

Xygalatas, D., Schjødt, U., Konvalinka, I., Jegindø, E.-M. E., Roepstorff, A., and Bulbulia, J. (2013b). Autobiographical memory in a fire-walking ritual. *J. Cogn. Cult.* 13, 1–16. doi: 10.1163/15685373-12342081

Zhong, C.-B., Bohns, V. K., and Gino, F. (2010). Good lamps are the best police: darkness increases dishonesty and self-interested behavior. *Psychol. Sci. a J. Am. Psychol. Soc. APS* 21, 311–314. doi: 10.1177/0956797609360754

Ziv, N. (2015). Music and compliance: can good music make us do bad things? *Psychol. Music.* doi: 10.1177/0305735615598855. [Epub ahead of print].

Ziv, N., Hoftman, M., and Geyer, M. (2012). Music and moral judgment: the effect of background music on the evaluation of ads promoting unethical behavior. *Psychol. Music* 40, 738–760. doi: 10.1177/0305735611406579

Zuckerman, P. (2007). "Atheism: contemporary rates and patterns," in *Cambridge Companion to Atheism,* ed M. Martin (Cambridge: Cambridge University Press), 46–67.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 March 2016
doi: 10.3389/fpsyg.2016.00371

Careful Cheating: People Cheat Groups Rather than Individuals

Amitai Amir[1], Tehila Kogut[2] and Yoella Bereby-Meyer[1]*

[1] Department of Psychology, Ben-Gurion University of the Negev, Beer Sheva, Israel, [2] Department of Education, Ben-Gurion University of the Negev, Beer Sheva, Israel

Cheating for material gain is a destructive phenomenon in any society. We examine the extent to which people care about the victims of their unethical behavior—be they a group of people or an individual—and whether they are sensitive to the degree of harm or cost that they cause to these victims. The results of three studies suggest that when a group (rather than a single individual) is the victim of one's behavior, the incidence of cheating increases only if the harm to the group is presented in global terms—such that the cheating might be justified by the relatively minor harm caused to each individual in the group (Studies #1 and #3). However, when the harm or cost to each individual in the group is made explicit, the tendency to cheat the group is no longer apparent and the tendency to cheat increases when the harm caused is minor—regardless of whether the victim is an individual or a group of people (Study #2). Individual differences in rational and intuitive thinking appear to play different roles in the decision to cheat different type of opponents: individual opponents seem to trigger the subject's intuitive thinking which restrains the urge to cheat, whereas groups of opponents seem to trigger the subject's rational mode of thinking which encourage cheating.

Keywords: ethics, morality, dishonesty, the singularity effect

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
David Reitter,
The Pennsylvania State University, USA
Enrico Rubaltelli,
University of Padua, Italy

****Correspondence:***
Tehila Kogut
kogut@bgu.ac.il

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *01 October 2015*
Accepted: *29 February 2016*
Published: *30 March 2016*

Citation:
Amir A, Kogut T and Bereby-Meyer Y (2016) Careful Cheating: People Cheat Groups Rather than Individuals. Front. Psychol. 7:371. doi: 10.3389/fpsyg.2016.00371

INTRODUCTION

Imagine a contractor who purchased all materials needed for a renovation job he is about to begin. When is he more likely to cheat about the cost of the materials and claim they costed more than what he actually paid: when the homeowner is a single person, or when the work is done for a condominium with several families?

In the present study, we set out to examine whether people care about who bears the consequences of their unethical behavior, and whether the degree of harm they cause when acting unethically plays any part in their decision. Specifically, we compare the action of deceiving a group of people as opposed to a single individual, and gage whether varying the degree of harm caused has any effect on people's behavior—either when presented in global terms, or when the respective harm to each individual in the group is stated explicitly.

People often engage in dishonest behaviors for material gain (Lewicki et al., 1997; Brief et al., 2001). However, research in the past decade consistently shows that people cheat only to the extent that they can maintain a self-concept of integrity (e.g., Mazar et al., 2008; Ayal and Gino, 2011). Thus, they are more likely to cheat when they feel they can justify their behavior, and the degree of cheating depends on the extent to which they can justify it to themselves (e.g., Shalvi et al., 2011).

However, little research has been done on the effect of the *identity* of the victim of unethical behavior on the tendency to cheat. Gneezy (2005) found that participants in his study were less likely to use deception to increase their payoffs at someone else's expense. However, in a competitive environment, where participants felt vulnerable in relation to their opponents, they *were* inclined to cheat, as this appears to have provided them with a strong justification to do so (Atanasov and Dana, 2011).

In many cases, it is easier to justify cheating a group by thinking that the harm caused would be distributed among several people, rather than borne by a single individual. Recently, Kesternich et al. (2014) analyzed distributional preferences in games in which decision makers choose the provision of a good that benefits a receiver and creates costs for a group of payers. They found that participants take into account the welfare of all parties and has concerns for efficiency. However, they attach similar weights to small and large groups of players alike, and tend to ignore large costs to the other party when these are shared by many individuals.

Cognitive research of people's perceptions of single individuals and groups (e.g., Hamilton and Sherman, 1996; Susskind et al., 1999) suggests that a single individual—in contrast to a group of individuals—is viewed as a psychologically coherent unit, which triggers a more extensive processing of information and active integration of the information in real time. As a result, people tend to be more emphatic in their assessment of an individual than of a group, and respond more quickly and confidently when asked to make a judgment about them. In contrast, the comparatively indistinct image of a group makes it easier for subjects to remain detached from it, and thereby easier to deceive for one's own benefit.

Research on pro-social decision making has confirmed that an individual victim elicits greater empathy and help than a group of victims in the same circumstances (e.g., Small and Loewenstein, 2003; Kogut and Ritov, 2005). Slovic (2007) suggests that this is because it is easier for people to put themselves in the shoes of one person than in the shoes of many. In addition, decisions about groups are expected to be more rational (i.e., take into account "objective" considerations), while decisions about individual victims are expected to be governed more by emotions (Kogut, 2011).

Given the global, more impersonal perception of the group, and the possibility of justifying one's behavior by the comparatively lesser harm inflicted on each individual in a group, we predicted that participants would tend to cheat a group of opponents more often than a single individual. Furthermore we examined whether informing participants of the specific harm or cost caused to each individual in a group of opponents would attenuate the tendency to cheat the group more than the individual.

To test these hypotheses, we conducted three studies, in which participants were asked to make private predictions of the outcomes of a series of coin tosses while playing against a single opponent or a group of four opponents, and receive payments according to the accuracy of their predictions (Zimerman et al., 2014). Since only the participants knew if their predictions were accurate, this task enabled them to earn more money by giving false reports of their predictions. Since we did not monitor the actual outcomes, we could not determine at the individual level whether or not a participant lied about their predictions, but we could compare their performance on an aggregate basis to that predicted by chance (see Batson et al., 1997; Shalvi et al., 2012; Fischbacher and Föllmi-Heusi, 2013).

In the present study, each correct prediction credited the player with a fixed amount of money—but unlike the above studies, these credits came at the expense of the earnings of an opponent, who was either a single individual or a group. In addition, while the amount earned by the players for each reported correct prediction remained constant in all conditions, the attendant cost to the opponent was varied (either *High* or *Low*). This enables us to examine exclusively the effect of the damage causes to the opponent on the tendency to cheat. In Study #1, the participant was informed only of the overall cost to the opponent group of his or her deception, without reference to the cost to each individual in it; in Study #2, they were informed of the cost incurred by each individual in the group; Study #3 included a direct comparison between the three conditions: a single opponent and the two group conditions that were examined in Studies #1 and #2, (i.e., with and without explicit information regarding the cost to each individual in the group).

STUDY #1

Method

One hundred and forty two undergraduate students (69 of whom were women, $M = 25.24$; $SD = 3.96$) were invited to participate in a short online experiment, and told that 10% of them would be randomly selected to earn money in accordance with their performance in the experiment.

The experiment involved a short task in which each participant was asked to toss a coin twenty consecutive times, after predicting the outcome in each case: for every correct prediction they made, they would earn a fixed amount of money, at the expense of their opponent's account (which would start with a particular amount). After each coin flip, they were asked to note the outcome on a separate screen, and indicate whether their prediction was correct or not.

Participants were randomly assigned to one of four experimental conditions of a 2 × 2 between-subject design involving two variables: *Cost to Opponent* (being either *high*—NIS 2.0 ∼ USD 0.50), or *low* (NIS 0.50 ∼ USD 0.13), and *Opponent Type* (an individual or a group of four people).

Although the Cost to Opponent varied between the *High* and *Low* conditions, the amount earned by the participant was constant in all instances—NIS 2. Thus, if they predicted all 20 tosses correctly, they could potentially earn as much as NIS 40 (∼US $10)—however, this would be at their opponent's expense. **Table 1** describes the four conditions, in which the Opponent Type (individual or group) and initial balance varied. As can be seen in the table, since the profit earned by the player remained constant in all conditions (2 NIS for each correct prediction) we kept the cost for the opponent group equal to the participants'

TABLE 1 | The four experimental conditions.

Cost to Opponent	Opponent: One individual	Opponent: Four individuals
Low (NIS 0.5)	Initial amount: NIS10 Deduction for each correct prediction: NIS 0.5	Initial balance: NIS 40 Deduction: NIS 2.0 total from the group as a whole (NIS 0.5 each)
High (NIS 2.0)	Initial amount: NIS 40 Deduction for each correct prediction: NIS 2.0	Initial balance: NIS 160 Deduction: NIS 8.0 from the group as a whole (NIS 2.0 each)

earnings, either at the group level (a total of NIS 2 per group, i.e., 0.5 for each individual in the group) or at the individual level (NIS 2 per each individual in the group, i.e., a total of NIS 8). The cost in the single opponent condition was adjusted to the costs per each individual in the group, and was either NIS 2 or NIS 0.5 (in the high and the low conditions, respectively). Participants in each condition were informed about the payoffs to the self and to their opponent precisely as reported in the table, without indication of the respective cost to each individual in the group condition.

Since research on the singularity effect highlighted the role of emotions and intuition in decisions that favor a specific target, we sought to examine the extent to which individual differences in Intuitive-Experiential and Analytical-Rational thinking are related to the decision to act unethically toward single opponents and toward groups. Hence, at the end of the experiment participants were asked to complete the short version of the Rational-Experiential Inventory questionnaire (REI; Epstein et al., 1996), comprising 10 items that gage individual differences between Intuitive-Experiential and Analytical-Rational thinking. (Cronbach's Alphas 0.86 and 0.82 for the Analytical-rational and the Intuitive-experiential scales, respectively; the correlation between the two scales was not significant ($r = 0.10$, $p = 0.30$).

Results and Discussion

Means (SDs) of the reported correct predictions are reported in **Table 2**. Overall, participants reported 11.17 correct predictions ($SD = 2.37$)—a significantly higher outcome than the expected chance performance rate (10); $t(141) = 5.88$, $p < 0.001$. This held true for both the *Individual Opponent* condition [10.82, $t(67) = 2.71$, $p = 0.009$] and the *Group Opponent* condition [11.49, $t(73) = 5.8$, $p < 0.001$].

To examine the role played by Opponent Type (individual or group), of the Cost to Opponent (*High* or *Low*), and of the two REI subscales (intuitive-experiential and analytical-rational thinking) in predicting the number of correct coin-tosses reported by the participants (0–20), we conducted a multiple regression analysis. As is recommended for binomial distributions, we performed an ARCSINE transformation of the proportion of correct predictions. The predictors included all four main effects, all two-way interactions and three-way interactions between these variables (see **Table 3**). The overall explained variance of the model was significant $F(11,130) = 1.92$, $p = 0.04$, $R^2 = 0.14$. The Opponent Type variable made a significant unique contribution to the model ($B = -3.53$, $t = -2.22$, $p = 0.028$). As expected, participants in the Group condition reported more correct predictions ($M = 11.49$) than those in the Individual Opponent condition ($M = 10.82$)—indicating a higher tendency to false reporting when the opponent was a group. In addition, both the interaction between Opponent Type and the Intuitive-Experiential subscale ($B = 0.96$, $t = 2.27$, $p = 0.025$) and between Opponent Type and the Analytical-Rational subscale ($B = 0.89$, $t = 2.21$, $p = 0.029$) were significant. These were plotted in **Figure 1** (right and left, respectively), as recommended by Aiken and West (1991) and Dawson and Richter (2006)[1]. As it demonstrates, when the opponent was

TABLE 2 | Mean (SDs) number of reported correct coin toss predictions, in each of the four conditions (Study #1).

Cost to Opponent	Individual Opponent	Group of 4 opponents	Total
Low (NIS 0.5)	$M = 10.72$	$M = 11.42$	$M = 11.09$
	($SD = 2.67$)	($SD = 2.29$)	($SD = 2.49$)
High (NIS 2.0)	$M = 10.91$	$M = 11.55$	$M = 11.25$
	($SD = 2.38$)	($SD = 2.14$)	($SD = 2.26$)
Total	$M = 10.82$	$M = 11.49$	
	($SD = 2.51$)	($SD = 2.20$)	

[1]According to Aiken and West (1991) in order to examine an interaction between two independent variables found in a regression analysis, the simple slopes of the regression should be plotted, using two meaningful points to anchor each line. Theoretically, one could choose any two values within the observed range of the DV to plot each line. It is most common to choose the mean, 1 SD below the mean, and 1 SD above the mean of a continuous variable.

TABLE 3 | The regression model – Study #1.

Model	Unstandardized Coefficients		*t*	Significant
	B	Standard Error		
(Constant)	1.982	0.300	6.613	0.000
Cost to Opponent	0.164	1.100	0.149	0.882
Opponent Type	−3.531	1.591	−2.219	0.028*
Analytical-Rational	0.034	0.071	0.476	0.635
Intuitive-Experiential	−0.128	0.052	−2.452	0.016*
Opponent * Rational	0.889	0.401	2.215	0.029*
Opponent * Intuitive	0.957	0.421	2.275	0.025*
Opponent * Cost	−0.023	0.096	−0.240	0.811
Cost * Rational	−0.144	0.286	−0.505	0.615
Cost * Intuitive	−0.100	0.284	−0.352	0.726
Cost * Rational * Intuitive	0.058	0.075	0.783	0.435
Opponent * Rational * Intuitive	−0.236	0.107	−2.200	0.030*

*$p \leq 0.05$.

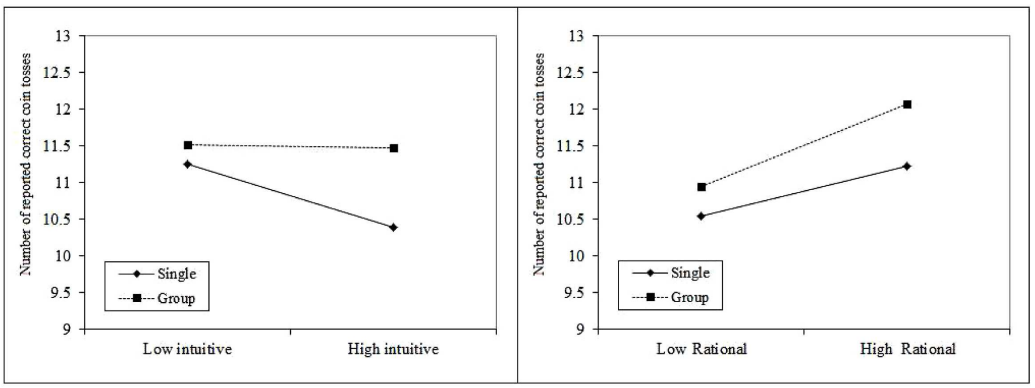

FIGURE 1 | Mean number of reported correct predictions, as a function of Opponent Type and the two REI sub-scales—plotted as recommended by Aiken and West (1991) and Dawson and Richter (2006), one SD below and one SD above the mean of each subscale in each of the Opponent Type conditions.

a single individual, Intuitive-Experiential ratings correlated negatively with correct predictions, the higher the Intuitive-Experiential tendency the lower the tendency to cheat. In addition, higher Analytical-Rational ratings were linked to increased incidence of cheating when the opponent was a group, and less so when the opponent was an individual. The three-way interaction between Opponent Type and the two Rational-Experiential scales was also significant ($B = -0.24$, $t = -2.20$, $p = 0.030$). No other significant interactions were found: in particular, the Cost to Opponent was not significant ($t = 0.15$, $p = 0.88$), nor did it significantly interact with any of the other variables.

The results of Study #1 indicate that people tend to cheat a group of opponents more often than an individual one—even when the harm or cost to each individual in the group is the same as that caused to the individual in the Single Opponent condition. Moreover, each condition appears to trigger a different mode of thinking: when faced with an individual opponent, the subject's Intuitive thinking tends to restrain their urge to cheat, whereas when faced with a group of opponents, the subject's Rational mode of thinking appears to encourage cheating. These results are in line with research on the *singularity effect*—namely, that a single opponent triggers a spontaneous emotional response in the subject that tends to result in decisions that are more favorable to the opponent (Kogut, 2009).

In the present study participants in the Group condition were informed of the cost of their deception to the group as a whole, without reference to the cost to each individual within the group. Thus, informing participants of the cost incurred by each individual in the opponent group may make people care more about each such individual, thereby making it more difficult for them to use the comparatively minor harm caused to each individual in the group as an excuse for cheating. To test this possibility, in Study #2 we replicated Study #1, but added information about the cost incurred by each group member in the Group Opponent conditions.

STUDY #2

Participants: One hundred and fifty two undergraduate students (81 of whom were women—$M = 24.8$, $SD = 1.89$), who were randomly assigned to one of the four experimental conditions of the same 2 × 2 between-subject design as in Study #1. Here too, they were invited to take part in a short online experiment, and told that 10% of them would be randomly selected to earn money according to their performance in the experiment. We used the same method as in Study #1—with one difference: in the Group Opponent condition, the participants were informed of the respective cost of their deception to each individual in the opponent group, as well as the total cost to the group as a whole.

Results and Discussion

The means (SDs) of the correct predictions reported in each condition are presented in **Table 4**. Overall, participants reported 12.21 correct predictions ($SD = 2.95$)—once again, significantly higher than the expected chance prediction rate (10); $t(151) = 9.24$, $p < 0.000$. This was true for both reports in the Single Opponent condition [12.32, $t(70) = 6.94$, $p = 0.001$] and in the Group condition [12.11, $t(80) = 6.19$, $p < 0.001$].

To examine the role of the two independent variables (Opponent Type and Cost to Opponent) in predicting the

TABLE 4 | Mean number of reported correct predictions (SD), in each of the four conditions (Study #2).

Cost to Opponent	Single Opponent	Group of 4	Total
Low	$M = 12.97$	$M = 12.47$	$M = 12.69$
(0.5 NIS)	($SD = 2.94$)	($SD = 3.11$)	($SD = 3.09$)
High	$M = 11.69$	$M = 11.71$	$M = 11.70$
(2 NIS)	($SD = 2.42$)	($SD = 3.00$)	($SD = 2.71$)
Total	$M = 12.32$	$M = 12.11$	
	($SD = 2.82$)	($SD = 3.01$)	

TABLE 5 | The three experimental conditions – Study #3.

A Single opponent	A global group	A detailed group
Initial amount: NIS 40 Deduction for each correct prediction: NIS 2.0	Initial balance: NIS 160 Deduction: NIS 8.0 from the group as a whole	Initial balance: NIS 160 Deduction: NIS 8.0 from the group (i.e., NIS 2.0 from each individual)

participants' reports, we conducted a 2 × 2 ANOVA on the transformed ARCSINE data of the proportion of the correct prediction of coin tosses as a function of the Opponent Type (individual or group) and the Cost to Opponent (High or Low). The results revealed no main effect for the Opponent Type $F(1,148) = 0.15$, (NS). However, the main effect for the Cost to Opponent approached significance [$F(1,148) = 3.72$, $p = 0.056$, $\eta_p^2 = 0.025$]—in that participants may have reported overall greater success in their predictions under the Low Cost condition ($M = 12.69$) than under the High Cost condition ($M = 11.70$). No significant interaction was found [$F(1,148) = 0.36$, NS].

The results of Study #2 indicate that when the actual cost to each individual in the opponent group is stated explicitly, participants likely care about the harm they may cause to others, irrespective of whether it is a single individual or a group. In these instances, the magnitude of harm or cost caused comes into play: when the harm or cost to each individual in the opponent group is minor, participants are inclined to cheat more often. In other words, informing the participant of the harm caused to each opponent group member appears to increase the participant's awareness of each individual in the group. It also appears to make it more difficult for the participants to use the relative minor harm to each individual in the group as a pretext for cheating (especially in the High Cost condition).

Taking the results of the two experiments together suggests that when the partner is a group, the extent of cheating depends on the way in which the information about the damage is presented: When the damage appears globally, without specifying the cost to each individual, people tend to cheat a group more than a single opponent; while, when the exact damage caused to each individual in the group is explicitly given, level of cheating groups and single opponents does not significantly differ. However, the two studies do not allow a direct comparison between the two groups (with and without explicit details on the extent of damage caused to each individual in the group), since different samples were examined, which differ in the overall extent of cheating. Hence, we conducted another study with three between subject groups: a single recipient, a global group (in which the damage caused to the group appears globally) and a detailed-group (in which the damage caused to each individual in the group is explicitly specified). In this study we kept the cost to each single opponent (whether an individual, or an individual in a group) constant (always two shekels, which is equal to the amount earned by the participant for each correct report).

STUDY #3

Participants: One hundred and nine undergraduate students (56 women—$M = 25.87$, $SD = 3.97$), who were randomly assigned to one of the three experimental between-subject conditions: (1) a single recipient, (2) a group of four recipients with information on the respective cost of a deception to the group as a whole (hereafter "Global-group"), and (3) a group of four recipients with information on the respective cost of a deception to each individual in the opponent group, as well as the total cost to the group as a whole (hereafter "Detailed-group"). Here too, participants were invited to take part in a short online experiment, and told that 10% of them would be randomly selected to earn money according to their performance in the experiment. The method was the same as in the high cost conditions in Studies #1 and #2 in the previous studies (see **Table 5**). In addition, as in Study #1, participants were asked to complete the short version of the Rational-Experiential Inventory questionnaire (REI; Epstein et al., 1996), at the end of the experiment (Cronbach's Alphas 0.80 and 0.85 for the Analytical-rational and the Intuitive-experiential scales, respectively; the correlation between the two scales was not significant ($r = -0.07$, $p = 0.41$).

Results and Discussion

Mean (SDs) number of reported correct coin toss predictions, in each of the three conditions are presented in **Table 6**. Overall, participants reported 11.06 correct predictions ($SD = 2.69$)—a significantly higher outcome than the expected chance performance rate (10); $t(108) = 4.12$, $p < 0.001$. However, in the *Single Opponent* condition, reported correct predictions were not significantly different from the ones expected by chance [10.62, $t(33) = 1.33$, $p = 0.191$]. The difference between reported correct predictions in the *Detailed-Group* condition and the expected outcome by chance (10.84) approached significance [$t(31) = 1.97$, $p < 0.058$]; while reports in the *Global-Group* condition were significantly different than the expected by chance [11.57, $t(42) = 3.61$, $p < 0.01$].

To examine the role played by Condition (Single, Global-group, and Detailed-group), and of the two REI subscales (Intuitive-experiential and Analytical-rational thinking) in predicting the number of correct coin-tosses reported by the participants (0–20), we conducted a multiple regression analysis on the ARCSINE transformation of the proportion of the correct prediction of coin tosses (see **Table 7**). Two dummy variables were created (Single and Detailed) using the

TABLE 6 | Mean (SDs) number of reported correct coin toss predictions, in each of the three conditions (Study #3).

Condition	Mean (*SD*)
Single Opponent	10.62 (2.70)
Detailed-*Group*	10.84 (2.37)
Global-*Group*	11.57 (2.88)

TABLE 7 | The regression model – Study #3.

Model	Unstandardized Coefficients		*t*	Significant
	B	Standard Error		
(Constant)	1.908	0.425	4.484	0.000
Single	0.549	0.272	2.014	0.047*
Detailed-group	0.563	0.338	1.667	0.099
Rational	−0.198	0.133	−1.489	0.140
Intuitive	−0.170	0.176	−0.967	0.336
Rational × Intuitive	0.120	0.056	2.161	0.033*
Single × Intuitive	−0.025	0.093	−0.269	0.788
Detailed-group × Intuitive	−0.247	0.118	−2.086	0.040*
Single × Rational	−0.245	0.080	−3.074	0.003**
Detailed-group × Rational	−0.023	0.087	−0.263	0.793

*$p \leq 0.05$; **$p \leq 0.01$.

Global group condition as the comparison group. Thus, the predictors included all four main effects (the Single and the Detailed-group dummies, the intuitive-experiential and the analytical-rational thinking scales and all two-way interactions between these variables. The overall explained variance of the model was significant $F(9,99) = 3.62$, $p = 0.001$, $R^2 = 24.8$. The contribution of the Single dummy – comparing the Global-group to the Single opponent conditions was significant ($B = 0.55$, $t = 2.01$, $p = 0.047$); such that participants in the Global-group reported more correct predictions (11.57) than participants in the Single opponent condition (10.62)—indicating a higher tendency to false reporting when the opponent was a Global-group, replicating the results of Study #1. The main effect of the Detailed-group dummy, comparing the Global-group to the Detailed-group, approached significance ($B = 0.56$, $t = 1.67$, $p = 0.099$); such that reported correct predictions were higher in the Global group than in the Detailed group (10.84). In addition, the interaction between the Single dummy and the Analytical-Rational subscale was significant ($B = -0.24$, $t = -3.07$, $p = 0.003$); such that higher Analytical-Rational ratings were linked to increased incidence of cheating only in the Global-group condition; while in the single opponent condition higher Analytical-Rational ratings were liked to a decrease in the number of reported correct predictions. Finally, the interaction between the Intuitive-experiential scale and the Detailed-dummy was significant ($B = -0.25$, $t = -2.09$, $p = 0.04$), showing that intuitive thinking is correlated with higher reports of correct predictions in the Global group condition, and with fewer reports of correct predictions in the Detailed-group condition.

In summary, the results of the third study support the conclusions of Studies #1 and #2 by showing that people tend to cheat a group more than a specific individual, mostly when the cost or harm to the group is presented in global terms. However, when the cost to each individual in the group is explicitly given, participants tend to cheat groups and individual opponents to a similar degree. The results also support the idea that higher analytical-rational thinking is related to the tendency to cheat a global group, replicating the results of Study #1. However, the results for the Intuitive-experiential scale were only partially consistent with the results of Study #1 by demonstrating a decrease in reported number of correct predictions in the Detailed-group condition (replicating the direction of results found in Study #1 for single opponents). However, in the global group condition the Intuitive-experiential scale was correlated with higher reports of correct predictions.

GENERAL DISCUSSION

Cheating is a destructive phenomenon in any society. When presented with an opportunity to profit by cheating, most people will do so—but only to a limited extent, to maintain their positive self-image as honest individuals (e.g., Mazar et al., 2008). However, when given a pretext for such behavior, they are more likely to cheat and to profit at the expense of others (e.g., Shalvi et al., 2011). Our results confirm that participants do cheat to some extent when faced with either a group of opponents or an individual one. However, if the consequent cost or harm to the group is presented in global terms, participants tend to cheat more often than when the opponent is an individual—perhaps because they imagine that the cost to each individual in the group is comparatively minor (Studies #1 and #3). Conversely, when they are explicitly told of the cost to each individual in the opponent group, participants tend to heed this and cheat groups and single opponents to the same degree (Studies #2 and #3). Furthermore, in such cases, participants are sensitive to the degree of harm or cost they cause, and tend to cheat more often when the harm they cause is less severe—regardless of whether their opponent is a single person or a group (Study #2). Since the amount of money earned from each correct prediction was the same in all studies, and is easily divisible by four (the number of people in the group), one might expect people to calculate the cost to each individual in the opponent group and show greater sensitivity as a result, even when this information is not explicitly stated. However, the results of our studies suggest that participants act only on the information given to them: when the cost of the deception to the group is presented in global terms, they appear to use the undefined cost to each individual opponent as a pretext for cheating.

As previously noted, research suggests that people perceive groups in a more global and impersonal fashion, which may induce a greater psychological distance and diminish their level of caring (Kogut and Ritov, 2005; Slovic, 2007). However, making the presence of each individual in a group more salient may increase a subject's level of concern for the group (Bartels and Burnett, 2011). This may have occurred in Studies #2 and #3, when participants were explicitly told of the respective loss that each member of the opponent group would incur as a result of their deception. This finding is in line with the *self-concept maintenance* model put forward by Mazar and Ariely (2006), which states that portraying unethical actions as more offensive may increase the internal cost of engaging in such actions, and discourage people from behaving dishonestly.

According to the standard economic model, individuals aim to maximize their own profit. A "rational" individual is therefore one who chooses the option that is expected to yield them the greatest profit (e.g., Hobbes and Macpherson, 1968; Smith, 1999). From this perspective, a person's decision to be honest depends only on the expected external benefits and costs to themselves (Lewicki, 1983; Hechter, 1990). The results of Studies #1 and #3—according to which individual differences in Analytical-Rational thinking predict a greater incidence of cheating—are in line with this model. However, this prediction was significant only when the opponent was a global group—not a single individual. The relatively detached perception of the group (compared with the perception of a single individual) allows for more "rational" thinking (as defined by the standard economic model), and appears to increase the incidence of cheating for monetary gain. Interestingly, when the opponent is a single individual (which possibly fosters greater perspective taking—Slovic, 2007), Intuitive-Experiential thinking tends to come into play, resulting in diminished cheating—possibly due to the subject's greater empathy toward the opponent. This pattern was also found for the Detailed-group, a group-setting that makes the individuals in the group more salient (Study #3). However, we did not find a replication for this direction in the single opponent condition in Study #3, and it is for future research to further explore this issue.

In many real-life situations, the precise harm caused by one's unethical behavior is not readily apparent. Our results suggest that in such cases people tend to cheat more often when their opponent is a group (as opposed to an individual)—perhaps on the assumption that the harm to each group member is minor compared with the harm that would have been caused to a single individual. Research has revealed various situations in which people fail to notice that their behavior violates their own moral standards (Gino et al., 2010; Bazerman et al., 2011; Schurr et al., 2012). Our research suggests that when considering a cost for a group of people, providing explicit information about the respective harm or cost that it would cause to each individual group member may reduce the incidence of such undesirable behavior, but only in cases when the relative cost to each group member is significant.

ETHICS STATEMENT

The research has been approved by the institutional review board of Ben-Gurion University of the Negev, Israel.

AUTHOR CONTRIBUTIONS

All coauthors designed the research. AA conducted first study. The second was conducted by a research assistant. All coauthors analyzed the data. All coauthors participated in writing the paper, which was led by TK.

FUNDING

This research was supported by Israel Science Foundation (ISF) Grant 1449/11.

ACKNOWLEDGMENT

This study is part of the first author's MA dissertation under the supervision of the second and third authors.

REFERENCES

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. London: Sage Publications.

Atanasov, P., and Dana, J. (2011). Leveling the playing field: dishonesty in the face of threat. *J. Econ. Psychol.* 32, 809–817. doi: 10.1016/j.joep.2011.07.006

Ayal, S., and Gino, F. (2011). "Honest rationales for dishonest behavior," in *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association).

Bartels, D. M., and Burnett, R. C. (2011). A group construal account of drop-in-the-bucket thinking in policy preference and moral judgment. *J. Exp. Soc. Psychol.* 47, 50–57. doi: 10.1016/j.jesp.2010.08.003

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., and Wilson, A. D. (1997). In a very different voice: unmasking moral hypocrisy. *J. Pers. Soc. Psychol.* 72, 1335–1348. doi: 10.1037/0022-3514.72.6.1335

Bazerman, M. H., Gino, F., Shu, L. L., and Tsay, C. J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emot. Rev.* 3, 290–292. doi: 10.1177/1754073911402370

Brief, A. P., Buttram, R. T., and Dukerich, J. M. (2001). "Collective corruption in the corporate world: toward a process model," in *Groups at Work: Theory and Research. Applied Social Research*, ed. M. E. Turner (Mahwah, NJ: Erlbaum), 471–499.

Dawson, J. F., and Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. *J. Appl. Psychol.* 91, 917–926. doi: 10.1037/0021-9010.91.4.917

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *J. Pers. Soc. Psychol.* 71, 390–405. doi: 10.1037/0022-3514.71.2.390

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Gino, F., Shu, L. L., and Bazerman, M. H. (2010). Nameless + Harmless = Blameless: when seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organ. Behav. Hum. Decis. Process.* 111, 102–115. doi: 10.1016/j.obhdp.2009.11.002

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Hamilton, D. L., and Sherman, S. J. (1996). Perceiving persons and groups. *Psychol. Rev.* 103, 336–355. doi: 10.1037/0033-295X.103.2.336

Hechter, M. (1990). The attainment of solidarity in intentional communities. *Ration. Soc.* 2, 142–155. doi: 10.1177/1043463190002002004

Hobbes, T., and Macpherson, C. B. (1968). *Leviathan; Edited with an Introduction by CB Macpherson*. Harmondsworth: Penguin Books.

Kesternich, I., Kosfeld, M., Schumacher, H., and Winter, J. (2014). *Us and Them: Distributional Preferences in Small and Large Groups, CESifo*

Working Paper Series No. 4657. Available at: http://ssrn.com/abstract=2407901

Kogut, T. (2009). Public decisions or private decisions? When the specific case guides public decisions. *J. Behav. Decis. Making* 22, 91–100. doi: 10.1002/bdm.613

Kogut, T. (2011). The role of perspective-taking and emotions in punishing identified and unidentified wrongdoers. *Cogn. Emot*. 25, 1491–1499. doi: 10.1080/02699931.2010.547563

Kogut, T., and Ritov, I. (2005). The "identified victim effect": an identified group, or just a single individual? *J. Behav. Decis. Making* 18, 157–167. doi: 10.1002/bdm.492

Lewicki, R. J. (1983). "Lying and deception: a behavioral model," *Negotiating in Organizations*, Vol. 68, eds M. H. Bazerman and R. J. Lewicki (Thousand Oaks, CA: Sage Publishing), 90.

Lewicki, R. J., Poland, T., Minton, J. W., and Sheppard, B. H. (1997). "Dishonesty as deviance: a typology of workplace dishonesty and contributing factors," in *Research on Negotiation in Organizations*, eds R. J. Lewicki, B. H. Sheppard and R. J. Bies (Greenwich, CT: JAI Press).

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Market. Res*. 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mazar, N., and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *J. Public Policy Market*. 25, 117–126. doi: 10.1509/jppm.25.1.117

Schurr, A., Ritov, I., Kareev, Y., and Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgm. Decis. Making* 7, 679–688.

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process*. 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychol. Sci*. 23, 1264–1270. doi: 10.1177/0956797612443835

Slovic, P. (2007). If I look at the mass I will never act: psychic numbing and genocide. *Judgm. Decis. Making* 2, 1–17.

Small, D. A., and Loewenstein, G. (2003). Helping a victim or helping the victim: altruism and identifiability. *J. Risk Uncertain*. 26, 5–16. doi: 10.1023/A:1022299422219

Smith, A. (1999). *The Wealth of Nations: Books IV and V*. London: Penguin Books.

Susskind, J., Maurer, K., Thakkar, V., Hamilton, D. L., and Sherman, J. W. (1999). Perceiving individuals and groups: expectancies, dispositional inferences, and causal attributions. *J. Pers. Soc. Psychol*. 76, 181–191. doi: 10.1037/0022-3514.76.2.181

Zimerman, L., Shalvi, S., and Bereby-Meyer, Y. (2014). Self-reported ethical risk taking tendencies predict actual dishonesty. *Judgm. Decis. Making* 9, 58–64.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 April 2016
doi: 10.3389/fpsyg.2016.00472

Are Some Countries More Honest than Others? Evidence from a Tax Compliance Experiment in Sweden and Italy

Giulia Andrighetto[1,2]*, *Nan Zhang*[3], *Stefania Ottone*[4], *Ferruccio Ponzano*[5], *John D'Attoma*[1] and *Sven Steinmo*[1,6]

[1] Robert Schumann Center for Advanced Studies, European University Institute, San Domenico di Fiesole, Fiesole, Italy, [2] National Council for Research, Institute of Cognitive Science and Technologies, Rome, Italy, [3] Department of Social and Political Sciences, European University Institute, San Domenico di Fiesole, Fiesole, Italy, [4] Department of Economics, Management and Statistics, University of Milano-Bicocca, Milano, Italy, [5] Department of Law, Politics, Economics and Social Sciences, University of Eastern Piedmont, Alessandria, Italy, [6] Department of Political Science, University of Colorado, Boulder, CO, USA

This study examines cultural differences in ordinary dishonesty between Italy and Sweden, two countries with different reputations for trustworthiness and probity. Exploiting a set of cross-cultural tax compliance experiments, we find that the average level of tax evasion (as a measure of ordinary dishonesty) does not differ significantly between Swedes and Italians. However, we also uncover differences in national "styles" of dishonesty. Specifically, while Swedes are more likely to be either completely honest or completely dishonest in their fiscal declarations, Italians are more prone to fudging (i.e., cheating by a small amount). We discuss the implications of these findings for the evolution and enforcement of honesty norms.

Keywords: tax compliance, ordinary dishonest behavior, fudging, cross-country comparison, social norms

OPEN ACCESS

Edited by:
Dan Ariely,
Duke University, USA

Reviewed by:
Ting Jiang,
University of Pennysalvania, USA
Stefania Bortolotti,
University of Cologne, Germany

***Correspondence:**
Giulia Andrighetto
giulia.andrighetto@gmail.com

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *17 July 2015*
Accepted: *17 March 2016*
Published: *07 April 2016*

Citation:
Andrighetto G, Zhang N, Ottone S, Ponzano F, D'Attoma J and Steinmo S (2016) Are Some Countries More Honest than Others? Evidence from a Tax Compliance Experiment in Sweden and Italy. Front. Psychol. 7:472. doi: 10.3389/fpsyg.2016.00472

INTRODUCTION

Ordinary dishonest behavior rarely attracts much attention. Seemingly innocuous practices such as avoiding VAT, double parking, cheating on an exam, and dodging fares on public transport tend to spread, often even in the wake of high-profile, sensationalized scandals. But while such everyday misdeeds may appear benign, taken together, they can result in vast societal damage (DePaulo et al., 1996; Ariely, 2008; Feldman, 2009; Ayal and Gino, 2011). In this study, we examine cross-national variation in individuals' willingness to engage in ordinary dishonest behavior, as measured by their tendency to underreport income for tax purposes.

The extent to which citizens engage in tax evasion and tax avoidance varies enormously across countries (Schneider and Enste, 2013). This is true even within European nations that share important features such as stable democratic institutions, developed economies, EU membership and broadly similar tax systems. Part of the reason underlying this cross-national variation relates to the efficiency of public institutions. Put simply, countries with efficient institutions (with stringent auditing and financial reporting standards) may be more effective at deterring tax evasion. At the same time, efficient institutions may encourage higher compliance because citizens feel that they are receiving something (i.e., high-quality public services) in return for their money (Levi, 1989; Smith and Stalans, 1991; Smith, 1992; Pommerehne et al., 1994; Edlund, 1999; Frey and Feld, 2002; Frey and Torgler, 2007; Torgler and Schneider, 2007; Cummings et al., 2009; Levi et al., 2009).

However, there is also reason to believe that variation in norms and culture plays an important role in explaining tax evasion. Consider two European countries that arguably lie at opposite ends

of the spectrum on tax compliance: Sweden and Italy[1]. Even setting aside differences in the institutional environment, substantial evidence suggests that norms of honesty may differ between these two countries. Specifically, Swedes think that honesty is a typical national trait (Daun, 1989), an assessment shared by other Europeans (Zetterberg, 1995)[2]. By contrast, Italy is ranked very low in terms of honesty amongst European countries, and even Italians themselves consider their compatriots to be less than trustworthy (Mackie, 2001). In fact, the Italian journalist and writer Giuseppe Prezzolini once described Italy as the "country of cunningness" (*paese dei furbi*), where people "worship cunningness so much that they even go so far as to admire those who use it against them" (Prezzolini, 1921).

To what extent can differences in norms and cultures of (dis)honesty explain cross-national variation in fiscal avoidance and evasion? To address this question, we report data from a tax compliance laboratory experiment conducted in Sweden and Italy in 2013/14[3]. Our experimental framework allows us to hold fiscal institutions constant, and thereby isolate the influence of national cultures on individuals' willingness to pay taxes. Given prevailing national stereotypes about norms of dishonesty, we expected that Italians would engage in greater fiscal evasion in the experiment, compared to Swedes.

To preview our basic findings, our experiment reveals, somewhat surprisingly, that *average* levels of tax evasion in Sweden and Italy do not differ significantly. Yet, we uncover country-specific styles of dishonesty. More specifically, we find that Italians engage more frequently in moderately dishonest behavior, or what Ariely (2012) refers to as "fudging." By contrast, Swedes are more likely to be perfectly honest in their behavior, but among those Swedes who do cheat, they are much more likely to cheat to the maximum extent possible. In the concluding section, we discuss some possible implications of Italians' greater tendency to fudge for the evolution and enforcement of honesty norms, with a particular eye toward explaining Italy's reputation as a "country of cunningness."

RELATED WORK

Several previous studies have attempted to evaluate cross-national variation in cheating and dishonesty using laboratory experiments. The results have been mixed. On the one hand, a number of studies have found that the propensity to engage in dishonest behavior does not diverge significantly across countries (Gneezy, 2005; Amir et al., 2008; Ariely, 2012; Pascual-Ezama et al., 2015; but see Dieckmann et al., 2015 for contradictory results). On the other hand, when honesty and dishonesty are measured in more real life domains (e.g., tax evasion and bribery scenarios) and framed language is used, systematic and predictable differences are observed across countries (Alm et al., 1995; Torgler, 2004; Bobek et al., 2007; Cummings et al., 2009; Barr and Serra, 2010).

Our study contributes to this literature in two ways. First, we suggest that national or cultural context can influence behavior in the lab under some conditions, but not necessarily others. This is because although honesty norms may differ across societies, normative considerations may have little effect on behavior if not first activated by situational cues in the decision context (Cialdini et al., 1990; Aarts and Dijksterhuis, 2003; Joly et al., 2008). For example, although the general norm in my society may be that "people should not lie," I could feel perfectly justified in lying to increase my payoffs in a lab experiment, *if* I believe that the operative norm *in that specific context* is to make as much money as possible. Given this, it is unsurprising that experiments using neutral language and context free tasks find little variation in dishonest behavior across countries (since the relevant country-specific norms remain dormant), whereas one finds variation when the specific context is made explicit and the corresponding norms are activated. For this reason, as we describe below, we designed our experiment to explicitly incorporate framed instructions in order to increase the salience of norms against tax evasion[4].

Secondly, we argue that a consideration of "average" country effects may obscure important variation in patterns of dishonesty. For example, suppose that researchers administer a matrix test to 20 participants, divided evenly between country A and country B. Suppose further that all 10 participants in country A cheat on 50% of the test questions, while in country B, 5 participants are completely honest, while 5 participants are completely dishonest. In this example, "average cheating" is identical across the two countries, but this average also masks important variation in the distribution (i.e., the extent and intensity) of dishonest behavior.

In relation to this last point, several studies have documented heterogeneity in degrees of dishonesty in experimental tasks (Gneezy et al., 2013). More specifically, one general finding emerging from the psychology literature is that, when given opportunities to be dishonest in everyday life, most people are willing to fudge—that is, to cheat "just a little bit" (Mazar et al., 2008; Gino et al., 2009; Ayal and Gino, 2011; Ariely, 2012). The attractiveness of fudging lies in its ability to reduce "ethical dissonance" by allowing people to recast their transgressions in

[1]One of the most obvious differences between these two countries is revealed in what is known as the "Tax Gap." The Tax Gap is a measure of the difference between revenues actually collected and taxes that would have been collected if all taxpayers had honestly reported their incomes. While it is difficult to precisely measure these gaps for obvious reasons, it is widely recognized that the Tax Gap in Sweden is approximately 8–9% of GDP (Slemrod, 2007), whereas in Italy it can reach as high as 25% to 30% (Santoro, 2010).

[2]In a recent YouGov poll, Northern Europeans perceived Sweden as the most honest nation in the EU (YouGov's Eurotrack Series, 2013).

[3]These experiments are part of a larger study on tax compliance behavior in five countries funded by the European Research Council.

[4]However, as noted by an anonymous reviewer, the use of framed instructions could introduce an experimenter demand effect: in particular, participants who wish to "look good" in front of the experimenters may behave more honestly. As we are interested in cross-national differences in behavior, this demand effect would be problematic for our analysis only if it also differs across countries. For example, Italians might care more about "looking good" than Swedes, and thus moderate the amount by which they cheat on their tax declarations in the experiment. However, we do not believe that this possibility poses a serious threat to the validity of our study. In particular, we were careful to ensure from the very beginning that participants had no knowledge that they were taking part in a cross-national comparative study. In other words, there is little reason for Italian (Swedish) participants to feel scrutinized just because they are Italian (Swedish). In addition, we use only native speakers (indeed, in Italy, only native dialect speakers) in each laboratory. This should lessen concerns that one needs to "look good" in front of foreign researchers.

a more benign light, and thereby reconcile dishonesty with the desire to maintain a positive moral self-image (Barkan et al., 2012).

In the context of the foregoing discussion, we are interested in examining how cross-national variation in social norms relating to tax evasion shapes both aggregate tax compliance as well as the tendency to engage in "fiscal fudging." Accordingly, both of these considerations—norm specificity and average vs. degrees of honesty—inform the design and analysis of the present study.

EXPERIMENTAL OVERVIEW

We report results from a tax experiment involving a total of 638 participants in Italy and Sweden (311 in Italy; 327 in Sweden), recruited in five different locations (Rome, Bologna, Milan, Stockholm and Gothenburg)[5] during the academic year 2013–2014[6]. The basic design of our experiment is similar to that used by Alm (1991), and aims to capture some essential features of the tax system used in many countries: (1) individuals earn real income, (2) they pay taxes on income voluntarily reported, (3) they face some chance that unreported taxes will be detected and penalized, and (4) the total taxes paid are used to provide a public good.

We describe our experimental protocols in detail below, but two features of our methodology are worth highlighting up front. First, our design explicitly provides a "context rich" setting in which tax language is used throughout. This feature is intended to ensure that participants' decisions in the lab reflect their experiences and social norms pertaining to the specific subject under study: taxation (Cummings et al., 2009). By contrast, the standard approach of using neutral language may encourage participants to perceive the decision problem at hand as a risky gamble (i.e., the extra income one earns from unreported taxes weighed against the probability of being caught and fined), as opposed to a tax compliance decision. An additional benefit of framing is that there is no ambiguity for participants about what constitutes honest behavior in the experiment. In other words, unlike in standard public good games in which participants may have different expectations about the appropriate amount of money to contribute, it is clear in the tax frame that the honest behavior is to declare the total amount earned.

Secondly, in our task, participants are not restricted to being either completely honest or completely dishonest, but instead, are allowed to report any amount (from 0 to 100%) of earned income. Thus, our task allows us to test whether Italians and Swedes differ in their tendency to fudge their taxes, an issue that has not been carefully investigated in previous work.

EXPERIMENTAL PROTOCOL

The experiment consisted of four stages, plus a post-experimental survey, and lasted 90 minutes on average. In this article, we report the findings of the first three stages of this experiment[7]. In all, we took great care to ensure that the participant pools were similar in each experimental location[8], and that the protocol was implemented in exactly the same manner in each country (Appendix Table 1 in the Supplementary Material displays descriptive statistics for each country sample, as well as the degree of similarity between Italian and Swedish participants)[9].

Each stage began with participants performing a 5 minute clerical task in which they copied random strings of letters and numbers from a sheet of paper onto an electronic form. Participants were paid 10 Experimental Currency Units (ECUs) for each line of text they correctly copied[10]. After the clerical task, participants were shown their earned income and asked to "report your income for tax purposes" under a variety of institutional scenarios (described below). Participants were not informed of how many scenarios would follow or what the specific content of each scenario might be.

In addition, participants were told that they would face a 5% probability of being audited in each scenario; if they underreported their income and were audited, they would pay a fine equal to twice the tax that they had avoided. Importantly, we revealed the results of any audits only at the end of the experiment, to avoid the possibility that being audited in one round would affect behavior in subsequent rounds. Moreover, throughout the experiment, participants had no knowledge of other participants' performance in the typing tasks or their tax reporting decisions. This ensured that individual choices did not reflect reciprocity or conditional cooperation.

In each of the three stages of the experiment, we manipulated fiscal rules relevant to different features of modern taxation systems, in order to elicit behavior under a range of institutional

[5]Replicating the experiment in multiple locations within each country provides us with greater confidence that we are not simply picking up "site-specific" effects, but rather cross-country differences in patterns of behavior. We chose these five locations specifically because they were the only active laboratories with suitable characteristics—i.e. with active participant pools drawn from different fields of study — that we could find in Sweden and Italy.

[6]Our experiments have been approved by the IRB Committee at the University of Colorado, Boulder, where the principal investigator holds a professorship. Our project has also been approved by the Ethics Council of the European Research Council, and the European University Institute Ethics committee. Finally, our work has also been authorized by all of the Italian and Swedish laboratories we have used, but we did not undergo a separate university-based IRB review in each case as these were not required by the universities in question. All participants signed a written consent form prior to taking part in the study.

[7]These three stages of the experiment encompass nine rounds of tax reporting (see Appendix Table 7 in the Supplementary Material for a summary). However, we report data from the first 8 rounds only. The 9th round involves donations to a real-world charity, and is not central to our research question. In addition, since the 9th round was the final round (and therefore, did not affect behavior in previous rounds), we have decided to exclude it from the analysis presented in this paper.

[8]Participants were all recruited using ORSEE (Greiner, 2004). In early versions of the experiment, the experimental tasks were programmed and conducted with zTree (Fischbacher, 2007), and the demographic information was collected through Qualtrics. Later in our project, we were able to integrate the experimental and survey portions of the study using our own web-based experimental software. A summary of the reporting rounds and a text version of the instructions (translated into English) are included in Appendix Table 7 in the Supplementary Material and Appendix Supplementary Information 8.

[9]We also had the protocol translated (double-blind) to ensure that the meanings of the words and phrases used were consistent across the countries.

[10]ECUs are converted into real currency at the end of the experiment. One ECU is worth €0.01 in Italy, and 0.60 SEK in Sweden. These exchange rates are chosen based on the average hourly pay rates in each country. The average earnings were 14.09 Euros in Italy and 187.60 SEK in Sweden.

contexts[11]. In stage 1, we altered the amount that participants received in return for the taxes that they collectively paid. In the first scenario (round 1) of stage 1, participants were simply told that the tax rate is 30%. There was no redistribution of tax revenues. In the second scenario (round 2), the tax rate remained 30%, but all tax revenues were placed in a "general fund" which was subsequently divided equally among all participants irrespective of how much each individual paid into the fund. In the third scenario (round 3), we again held the tax rate at 30%, but all tax revenues in the general fund were *doubled* and then redistributed equally to all participants, regardless of how much each participant had individually paid into the fund. In each round (before they were asked to report their incomes), subjects were given multiple specific examples demonstrating the rules in each scenario under a series of hypothetical decisions (see Appendix Supplementary Information 8 in the Supplementary Material); they were also reminded of the 5% probability of being audited, as well as of the fine they would have to pay should the audit detect any under-reporting.

In stage 2, we held redistribution constant and varied the tax rates. In the first scenario of stage 2 (round 4), we asked participants to report their income under a tax rate of 10%. In the second scenario (round 5), the tax rate was increased to 30%, and in the third scenario (round 6), the tax rate was again increased to 50%. In all three rounds of stage 2, we held the audit rate (5%), fines (2x underreported income) and the rules for redistribution (tax revenues doubled and then redistributed) constant.

[11] We also considered randomly ordering the scenarios to control for order effects. However, we decided that this option was unnecessary because our central concern is not to evaluate the effects of institutional changes, but rather to examine how people in different countries would respond to the same institutional scenarios.

Finally, in stage 3, we presented scenarios with two different types of progressive taxation schemes. In round 7, the top 10% of income earners (as defined by self reported income) faced a 50% tax rate; participants in the bottom 10% of reported incomes faced a 10% rate; finally, the middle 80% of reported income earners faced a 30% rate. By contrast, in round 8, we introduced a *marginal tax rate system* (similar to the real tax systems operating in Italy and Sweden). In this case, all subjects paid a 10% tax on the first 50 ECUs of reported income, a 30% tax on reported income between 51 and 100 ECUs, and a 50% tax on all reported income above 100 ECUs. In both progressive taxation rounds, all tax revenues were doubled and then redistributed, and we held the audit rate constant at 5%. Once again, subjects were given explicit examples to ensure their understanding of the rules.

RESULTS

Average Compliance Rate

Despite the intrinsic social dilemma structure of the tax scenario that makes evasion the optimal strategy, we find that the level of compliance far exceeded the level predicted by expected utility theory (Allingham and Sandmo, 1972; Yitzhaki, 1974) in both countries and in all rounds. This result is consistent with previous research on tax compliance and public goods (Ledyard, 1995; Bosco and Mittone, 1997; Cummings et al., 2009; Alm, 2012). Pooling both countries, we observe that individuals were mostly honest, reporting on average 64.9% of total income.

Additionally, we observe that the reporting rate varied according to the specific scenarios presented in each round. **Figure 1** shows the average percentage of earned income that was reported in each of the eight rounds, broken down between Swedish and Italian participants. The vertical axis displays

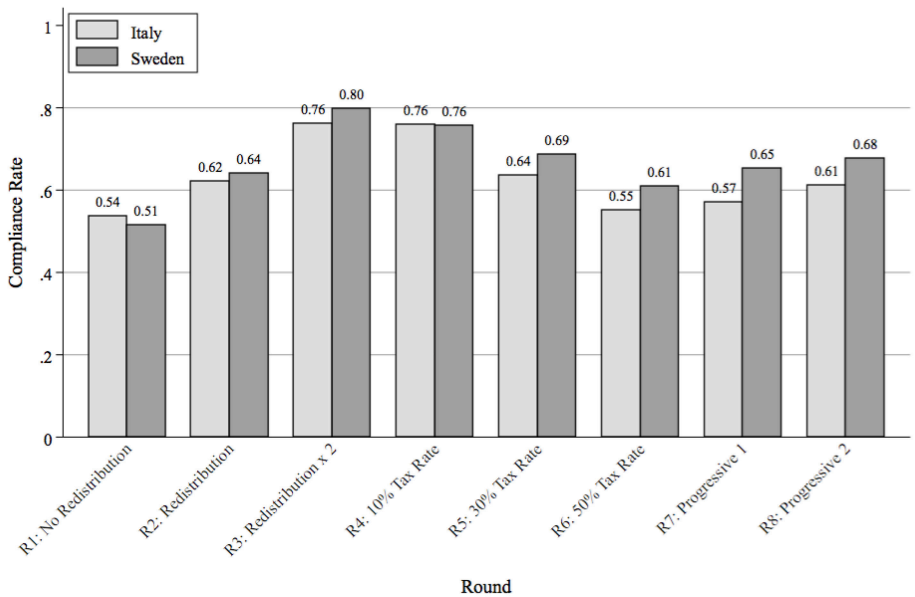

FIGURE 1 | Average compliance rate divided by round and by country.

the average tax compliance rate, defined as the percentage of total earned income that is truthfully declared in each round. Comparing rounds 1 through 3, we see that compliance responds positively to the efficiency of redistribution: individuals were willing to declare more when they knew that tax revenues produced more public goods. Secondly, individuals responded to higher tax rates by evading their fiscal obligations: compliance falls moving from rounds 4 through 6. These results are in line with previous experimental studies on tax compliance (Alm et al., 1992; Bosco and Mittone, 1997; Torgler, 2002; Blackwell, 2007; Alm, 2012), providing us with some assurance about the validity of our experimental design.

Turning to the cross-country variation in average compliance rates, although we predicted that Swedes would comply more on average than Italians, we do *not* document significant differences across countries. Pooling across all 8 rounds of the experiment, Italians reported 63.1% of their earned income (s.e. = 1.8%), as compared to Swedes who reported 66.6% (s.e. = 1.9%), and the cross-country difference is only 3.5% (*t*-test s.e. = 2.6%, $p = 0.182$). We run several additional tests to assess the robustness of this result. First, we check whether different locations within each country can indeed be pooled to estimate a larger "country" effect. To do so, in Models 1 and 2 of Appendix Table 2 in the Supplementary Material, we estimate individual-level tobit models for the average compliance rate (pooled across all 8 rounds) with site-specific dummy variables, separately for Italy and Sweden[12]. We also cluster standard errors by experimental session. We find no statistically significant within-country variability, suggesting that the results from different locations can indeed be pooled.

Next, we put data from both countries together, and estimate the effect of a dummy variable for Italian participants on the average compliance rate, controlling for a host of individual-level characteristics including gender, age, previous participation in experiments, economics training, earnings in the clerical task, and beliefs about the honesty of other participants. In an alternative specification, we also add fixed effects for the individual treatment round. The inclusion of covariates in Models 3 and 4 of Appendix Table 2 in the Supplementary Material allows us to examine individual-level correlates of tax evasion and dishonesty. We observe that the average compliance rate is lower amongst men, and amongst younger participants (although this latter result is less robust), which is consistent with previous research (Hasseldine, 1999; Lewis et al., 2009; Torgler and Valev, 2010). Risk aversion is also correlated with higher average compliance[13]. In addition, in line with previous work, we find a positive correlation between economics training and lower average compliance (Marwell and Ames, 1981; Carter and Irons, 1991; Cullis et al., 2006; Lewis et al., 2009). Finally, we control for participants' beliefs about the behavior of others in the experiment[14]. Individuals who believed that others reported less also reported less themselves (Fischbacher et al., 2001).

Importantly, the inclusion of these covariates does not change our overall conclusions regarding cross-country differences in the average compliance rate. As shown in Models 3 and 4, the coefficient on the Italy dummy is never statistically significant. These additional results confirm our initial findings reported above: regardless of the controls and model specification we employ, we do not find any significant differences in average compliance rates between the two countries[15].

Patterns of Compliance and Dishonesty

Although an analysis of the *average compliance rate* does not support prevailing national stereotypes that Swedes are more honest than Italians, a closer analysis of the *distribution* of compliance decisions yields some interesting cross-national differences. In particular, a statistic like the average compliance rate does not allow us to distinguish between three different decisions: complete compliance (i.e., the decision to declare 100% of earned income), complete evasion (i.e., the decision to declare 0% of earned income), and partial compliance or "fudging" (i.e., the decision to declare more than 0, but less than the total; see also Mazar et al., 2008 for a similar analysis).

These distinctions are shown in **Figure 2**, which displays the distribution of participants' reported incomes (pooled across all 8 rounds). The x-axis breaks down the distribution of reported incomes into the following bins: [0%, (1–10%), (11–20%)…(91–99%), 100%], and the y-axis displays the percentage of participants in each country falling into each bin. We observe that Swedes tended to concentrate in the extreme bins (0% and 100%), while the distribution is more uniform amongst Italians.

To more precisely operationalize these patterns, we define the following three "types" of participants:

- Honest Type: declares 100% of earned income across all 8 rounds.
- Dishonest Type: declares 0% of earned income across all 8 rounds.
- Fudging Type: everyone else.

Next, we compare the distribution of types across Italy and Sweden. We find more Honest Types in Sweden compared to Italy (25.7% in Sweden vs. 14.8% in Italy; Schlag z-test $p < 0.001$), but also more Dishonest Types (8.9% in Sweden vs. 5.1% in Italy; Schlag z-test $p = 0.066$). By contrast, significantly more Italians are classified as Fudging Types (80% in Italy vs. 65% in Sweden;

[12]The number of observations changes once we include demographic covariates in our regression models. This is because in early versions of the experiment, the experimental tasks were implemented in zTree, while the demographic information was collected separately using Qualtrics. This necessitated participants entering their anonymous participant-IDs twice: once into zTree, and once again into Qualtrics. Because some participants accidentally entered different participant-IDs into the two systems, we were unable to match their experimental decisions with their demographic data. This problem was fixed in later versions of the experiment, once we switched to our own web-based experimental software.

[13]We measured risk using a survey item that asks subjects to rank themselves on a 10-point scale, with 1 signifying a person who "normally tries to avoid taking risks" and 10 signifying someone who is "completely willing to take risks." Answers have been standardized to have mean = 0 and s.d. = 1.

[14]We measured participants' perceptions a survey item which asks subjects whether they thought other participants in the experiment reported (a) their entire earned incomes, (b) less than their entire earned incomes, or (c) much less than their entire earned incomes. In our regressions, we use (b) as our baseline category.

[15]As a further robustness check, we compared country-level differences in average compliance rates separately for each individual round of the experiment. In 6 out of the 8 rounds, we found no statistically significant differences (see Appendix Table 3 in the Supplementary Material).

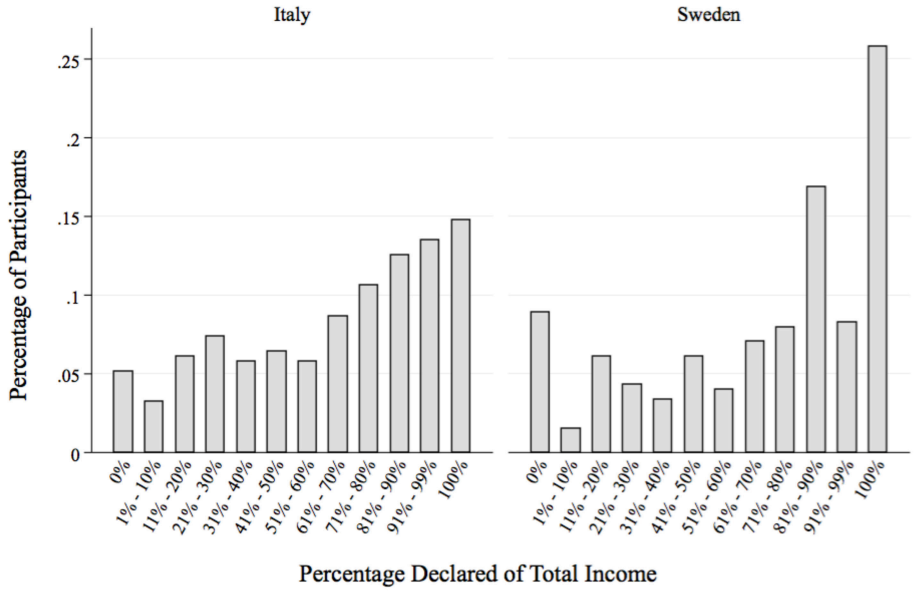

FIGURE 2 | Distribution of individual compliance rates.

Schlag z-test $p < 0.001$). In other words, Swedish participants displayed more clear-cut behaviors: Swedes cheat less frequently, but when they cheat, they are likely to do so completely. By contrast, Italians cheat more habitually, but the intensity of their cheating is more restrained: they hold back from "cheating all they way."

Interestingly, we also find that, compared to Dishonest Types, Fudging Types are also more likely to deceive (themselves) about their behavior during the experiment. In particular, in our post-experimental survey, participants were asked to indicate how much of their total earnings they *themselves* reported during the experiment: 18% of Fudging Types indicating that they reported their total income, while no Dishonest Types lied. This last finding nicely fits with evidence from social psychological research showing that individuals choose fudging strategies to maintain a positive moral reputation and self-image (Ayal and Gino, 2011; Ariely, 2012).

To check the robustness of these results, we conduct an additional battery of tests. First, as before, we verify that results from separate locations within countries can indeed be pooled (Models 1 and 2 of Appendix Table 4 in the Supplementary Material)[16]. Next, we estimate probit models of the probability of being a Fudging Type, conditional upon individual-level covariates and round fixed effects (Models 3 and 4 of Appendix Table 4 in the Supplementary Material). In all specifications, Italians were approximately 10% more likely to fudge, compared to Swedes. Here, we also find that individuals who believed that others behaved honestly in the experiment were significantly less likely to fudge[17,18].

In summary, although the average *level* of dishonesty does not differ across the two countries, a closer examination of the data reveals a cross-national difference in *patterns* of dishonesty. Simply put: Italians are more prone to "fudging" than Swedes.

DISCUSSION AND CONCLUDING REMARKS

Our results indicate that when Italians and Swedes face a tax compliance scenario consisting of a transparent tax system, efficient redistributive regime, and clear audit rules and penalties, the average level of honesty is relatively high in both countries. This result does not bear out our initial expectations based on national stereotypes, where we predicted a greater level of honesty in Sweden compared to Italy. However, we also identify an interesting cross-country difference that may shed light on our understanding of why these stereotypes emerge. In particular, we find country-specific styles of dishonesty, with Italians engaging more frequently in fudging, while Swedes were more likely to be both perfectly honest and perfectly dishonest. In this concluding section, we offer some conjectures linking this result to the

[16]The percentage of Fudging Types in all Italian locations is higher than in all Swedish locations (83% in Milan, 74% in Bologna and 84% in Rome vs. 67% in Stockholm and 62% in Gothenburg). Running an "empty" random-effects model, we find that the variance within countries is about half the size of the variance across countries.

[17]We also check whether our results are sensitive to the definition of fudging we employ. Specifically, we alternatively redefine Fudging Types as those who reported (a) more than $2/3^{rds}$ of their income, (b) between $1/3^{rd}$ and $2/3^{rds}$ of total income, and (c) less than $1/3^{rd}$ of total income. Overall, as shown in Appendix Table 5 in the Supplementary Material, we find that regardless of the definition of Fudging Type, Italians were more likely fudge.

[18]We also checked for cross-country differences in the distribution of types separately for each individual round. We find that in all 8 rounds of the experiment, Italians were significantly more likely to fudge than Swedes (See Appendix Table 6 in the Supplementary Material).

development and perpetuation of national stereotypes about honesty and dishonesty in Sweden and Italy.

In particular, we argue that when ordinary dishonesty takes on the form of fudging, this behavior may be particularly difficult to control and eradicate. Part of the reason stems from the fact that fudging introduces a degree of moral ambiguity in judging the wrongfulness of a particular action. As discussed in Ayal and Gino (2011), when the categorization of a behavior is malleable rather than clear-cut, people are more likely to conceptualize their own actions in acceptable terms. This benevolent interpretation of dishonest behavior helps to reduce any dissonance that may result from the tension between unethical conduct and the desire to maintain a moral self-image (Baumeister, 1998; Schweitzer and Hsee, 2002). Fudging thus provides individuals with greater moral license to indulge in (moderate) wrongdoing.

In addition, in the presence of widespread fudging, it may be difficult for third parties to enforce honesty norms. In particular, when there is uncertainty about what is right or wrong, punishment becomes more risky, since enforcement may generate counter-punishment (also from third-party observers) who do not recognize the legitimacy of the punisher (Herrmann et al., 2008; Strimling and Eriksson, 2014). As such, tolerance for (moderate) wrongdoing rises.

Given the difficulties that fudging poses for both self-regulation and peer-regulation of dishonest behavior, ordinary dishonesty tends to spread. This may explain why Italians have such a widespread reputation for cunningness, as they are observed both to engage in ubiquitous small acts of dishonesty, and to tolerate and even justify dishonesty on the part of others. By contrast, Swedes' relatively clear-cut behaviors may facilitate both self-regulation (as it is more difficult to self-justify gross dishonesty) and social control.

Efforts to raise the moral standard of society in the presence of fudging may thus require actions that (a) increase awareness of the negative effects of apparently benign behaviors, and (b) support norms enforcers who insist on absolute honesty. In future work, we propose to use agent-based modeling and additional experiments to explore the dynamics of fudging, its social effects, and the effectiveness of policy interventions to foster greater public integrity.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: GA, SO, FP, and SS. Performed the experiments: GA, SO, FP, SS, and NZ. Analyzed the data: NZ, SO, and JD. Wrote the paper: GA.

FUNDING

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement n. (295675). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00472

REFERENCES

Aarts, H., and Dijksterhuis, A. (2003). The silence of the library: environment, situational norm, and social behavior. *J. Pers. Soc. Psychol.* 84, 18–28. doi: 10.1037/0022-3514.84.1.18

Allingham, M., and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *J. Public Econ.* 1, 323–338. doi: 10.1016/0047-2727(72)90010-2

Alm, J. (1991). A perspective on the experimental analysis of taxpayer reporting. *Account. Rev.* 66, 577–593.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *Int. Tax Public Finance* 19, 54–77. doi: 10.1007/s10797-011-9171-2

Alm, J., Jackson, B. R., and McKee, M. (1992). Estimating the determinants of taxpayer compliance with experimental data. *Natl. Tax J.* 45, 107–114.

Alm, J., Sanchez, I., and De Juan, A. (1995). Economic and noneconomic factors in tax compliance. *Kyklos* 48, 3–18. doi: 10.1111/j.1467-6435.1995.tb02312.x

Amir, O., Ariely, D., and Mazar, N. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–634. doi: 10.1509/jmkr.45.6.633

Ariely, D. (2008). *Predictably Irrational*. New York, NY: HarperCollins.

Ariely, D. (2012). *The (Honest) Truth About Dishonesty*: *How We Lie to Everyone–Especially Ourselves*. New York, NY: HarperCollins

Ayal, S., and Gino, F. (2011). "Honest rationales for dishonest behavior," in *The Social Psychology of Morality: Exploring the Causes of Good and Evil. Herzliya Series on Personality and Social Psychology*, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 149–166.

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *J. Exp. Psychol. Gen.* 141, 757–773. doi: 10.1037/a0027588

Barr, A., and Serra, D. (2010). Corruption and culture: an experimental analysis. *J. Public Econ.* 94, 862–869. doi: 10.1016/j.jpubeco.2010.07.006

Baumeister, R. F. (1998). "The self," in *Handbook of Social Psychology,* eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: McGraw-Hill), 680–740.

Blackwell, C. (2007). "A meta-analysis of tax compliance experiments," in *Working Paper Series, at AYSPS, 0724 International Center for Public Policy*. International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Bobek, D. D., Roberts, R. W., and Sweeney, J. T. (2007). The social norms of tax compliance: evidence from Australia, Singapore, and the United States. *J. Bus. Ethics* 7, 49–64. doi: 10.1007/s10551-006-9219-x

Bosco, L., and Mittone, L. (1997). Tax evasion and moral constraints: some experimental evidence. *Kyklos* 50, 297–324.

Carter, J., and Irons, M. (1991). Are economists different, and if so, why? *J. Econ. Perspect.* 5, 171–177. doi: 10.1257/jep.5.2.171

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct—recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015

Cullis, J., Jones, P., and Lewis, A. (2006). Tax framing, instrumentality and individual differences: are there two different cultures? *J. Econ. Psychol.* 27, 304–320. doi: 10.1016/j.joep.2005.07.003

Cummings, R., Martinez-Vazquez, J., McKee, M., and Torgler, B. (2009). Tax morale affects tax compliance: evidence from surveys and an artefactual field experiment. *J. Econ. Behav. Organ.* 70, 447–457. doi: 10.1016/j.jebo.2008.02.010

Daun, A. (1989). *Swedish Mentality*. University Park, TX: The Pennsylvania University Press.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

Dieckmann, A., Fischbacher, U., Grimm, V., Unfried, M., Utikal, V., and Valmasoni, L. (2015). *Trust and Beliefs among Europeans: Cross-Country Evidence on Perceptions and Behavior*. Institut fur Wirtschaftspolitik und Quantitative Wirtschaftsforschung. Discussion Papers 04/2015.

Edlund, J. (1999). Trust in government and welfare regimes: attitudes to redistribution and financial cheating in the USA and Norway. *Eur. J. Polit. Res.* 35, 341–370. doi: 10.1111/1475-6765.00452

Feldman, R. (2009). *The Liar in Your Life*. London: Virgin books.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* 71, 397–404. doi: 10.1016/S0165-1765(01)00394-9

Frey, B., and Feld, L. (2002). *Deterrence and Morale in Taxation: An Empirical Analysis*. Technical report CESifo working paper Number 760.

Frey, B., and Torgler, B. (2007). Tax morale and conditional cooperation. *J. Comp. Econ.* 35, 136–159. doi: 10.1016/j.jce.2006.10.006

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Gneezy, U., Rockenbach, B., and Serra-García, M. (2013). Measuring lying aversion. *J. Econ. Behav. Organ.* 93, 293–300. doi: 10.1016/j.jebo.2013.03.025

Greiner, B. (2004). *The Online Recruitment System ORSEE 2.0-A Guide for the Organization of Experiments in Economics.* University of Cologne, Working Paper Series in Economics, 63–104.

Hasseldine, J. (1999). Gender differences in tax compliance. *Asia Pac. J. Taxation* 3, 73–89.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science* 319, 1362–1367. doi: 10.1126/science.1153808

Joly, J. F., Stapel, D. A., and Lindenberg, S. (2008). Silence and table manners: when environments activate norms. *Pers. Soc. Psychol. Bull.* 34, 1047–1056. doi: 10.1177/0146167208318401

Ledyard, J. O. (1995). "Public goods: a survey of experimental research," in *Handbook of Experimental Economics,* eds J. H. Kagel and A. E. Roth (Princeton, NJ: Princeton University Press), 111–194.

Levi, M. (1989). *Of Rule and Revenue*. Berkeley, CA: University of California Press.

Levi, M., Sacks, A., and Tyler, T. (2009). Conceptualizing legitimacy, measuring legitimating beliefs. *Am. Behav. Sci.* 53, 354–375. doi: 10.1177/0002764209338797

Lewis, A., Carrera, S., Cullis, J., and Jones, P. (2009). Individual, cognitive and cultural differences in tax compliance: UK and Italy compared. *J. Econ. Psychol.* 30, 431–445. doi: 10.1016/j.joep.2008.11.002

Mackie, G. (2001). "Patterns of social trust in western europe and their genesis," in *Trust in Society,* ed K. Cook (New York, NY: Russell Sage Foundation), 245–282.

Marwell, G., and Ames, R. (1981). Economists free ride, does anyone else? *J. Public Econ.* 15, 295–310. doi: 10.1016/0047-2727(81)90013-X

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Market. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Vesztag, R., Gil-Gomez de Liano, B., et al. (2015). Context dependent cheating: experimental evidence from 16 countries. *J. Econ. Behav. Organ.* 116, 379–386. doi: 10.1016/j.jebo.2015.04.020

Pommerehne, W., Hart, A., and Frey, B. (1994). Tax morale, tax evasion and the choice of policy instruments in different political systems. *Public Finance* 49, 52–69.

Prezzolini, G. (1921). *Codice Della Vita Italiana.* Florence: La Voce.

Santoro, A. (2010). *L'Evasione Fiscale*. Bologna: Mulino.

Schneider, F., and Enste, D. (2013). *The Shadow Economy: An International Survey*. New York, NY: Cambridge University Press.

Schweitzer, M. E., and Hsee, C. K. (2002). Stretching the truth: elastic justification and motivated communication of uncertain information. *J. Risk Uncertain.* 25, 185–201. doi: 10.1023/A:1020647814263

Slemrod, J. (2007). Cheating ourselves: the economics of tax evasion. *J. Econ. Perspect.* 21, 25–48. doi: 10.1257/jep.21.1.25

Smith, K. (1992). "Reciprocity and fairness: positive incentives for tax compliance," in *Why People Pay Taxes. Tax Compliance and Enforcement,* ed J. Slemroad, J. (Ann Arbor, MI: University of Michigan Press), 223–250.

Smith, K., and Stalans, L. (1991). Encouraging tax compliance with positive incentives: a conceptual framework and research directions. *Law Policy* 13, 35–53. doi: 10.1111/j.1467-9930.1991.tb00056.x

Strimling, P., and Eriksson, K. (2014). "Regulating the regulation: norms about how people may punish each other," in *Social Dilemmas: Punishment and Rewards,* eds P. van Lange, T. Yamagishi, and B. Rockenbach (Oxford, UK: Oxford University Press), 52–69.

Torgler, B. (2002). Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. *J. Econ. Surv.* 16, 657–683. doi: 10.1111/1467-6419.00185

Torgler, B. (2004). Cross-culture comparison of tax morale and tax compliance: evidence from Costa Rica and Switzerland. *Int. J. Comp. Sociol.* 45, 17–43. doi: 10.1177/0020715204048309

Torgler, B., and Schneider, F. (2007). What shapes attitudes toward paying taxes? Evidence from multicultural european countries. *Soc. Sci. Q.* 88, 443–470. doi: 10.1111/j.1540-6237.2007.00466.x

Torgler, B., and Valev, N. (2010). Gender and public attitudes toward corruption and tax evasion. *Contemp. Econ. Policy* 28, 554–568. doi: 10.1111/j.1465-7287.2009.00188.x

Yitzhaki, S. (1974). A note on income tax evasion: a theoretical analysis. *J. Public Econ.* 3, 201–202. doi: 10.1016/0047-2727(74)90037-1

YouGov Eurotrack (2013). *Results 290513 Honesty*. London: Europe EuroTrack Life.

Zetterberg, H. L. (1995). "Valuescope: a three-dimensional value system," in *European Advances in Consumer Research,* Vol. 2, ed H. E. Flemming (Provo, UT: Association for Consumer Research), 163–171.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 February 2016
doi: 10.3389/fpsyg.2016.00085

What Deters Crime? Comparing the Effectiveness of Legal, Social, and Internal Sanctions Across Countries

Heather Mann[1], Ximena Garcia-Rada[2], Lars Hornuf[3] and Juan Tafurt[4]*

[1] Department of Psychology and Neuroscience, Duke University, Durham, NC, USA, [2] Social Science Research Institute, Duke University, Durham, NC, USA, [3] Department of Economics and IAAEU, University of Trier, Trier, Germany, [4] AG3 Consultores, Bogota, Colombia

The question of what deters crime is of both theoretical and practical interest. The present paper focuses on what factors deter minor, non-violent crimes, i.e., dishonest actions that violate the law. Much research has been devoted to testing the effectiveness of legal sanctions on crime, while newer models also include social sanctions (judgment of friends or family) and internal sanctions (feelings of guilt). Existing research suggests that both internal sanctions and, to a lesser extent, legal sanctions deter crime, but it is unclear whether this pattern is unique to Western countries or robust across cultures. We administered a survey study to participants in China, Colombia, Germany, Portugal, and USA, five countries from distinct cultural regions of the world. Participants were asked to report the likelihood of engaging in seven dishonest and illegal actions, and were asked to indicate the probability and severity of consequences for legal, friend, family, and internal sanctions. Results indicated that across countries, internal sanctions had the strongest deterrent effects on crime. The deterrent effects of legal sanctions were weaker and varied across countries. Furthermore, the deterrent effects of legal sanctions were strongest when internal sanctions were lax. Unexpectedly, social sanctions were positively related to likelihood of engaging in crime. Taken together, these results suggest that the relative strengths of legal and internal sanctions are robust across cultures and dishonest actions.

Keywords: dishonesty, crime, cheating, cross-cultural, deterrence theory, deterrence

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Stefan T. Trautmann,
University of Heidelberg and Tilburg University, Germany
Ori Weisel,
University of Nottingham, UK

****Correspondence:***
Heather Mann
heather.mann@gmail.com

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *02 July 2015*
Accepted: *15 January 2016*
Published: *08 February 2016*

Citation:
Mann H, Garcia-Rada X, Hornuf L and Tafurt J (2016) What Deters Crime? Comparing the Effectiveness of Legal, Social, and Internal Sanctions Across Countries. Front. Psychol. 7:85. doi: 10.3389/fpsyg.2016.00085

INTRODUCTION

The question of what deters crime is of interest to social science researchers and policy-makers alike. Are decisions to engage in crime influenced by the threat of legal consequences? Are they influenced by threats of judgment from friends or family? Are they influenced by the potential for internal feelings of guilt? These questions are relevant to any society, as dishonesty can be extremely costly. For example, it is estimated that for most countries, losses due to tax evasion are greater than the total amount spent on healthcare (The Tax Justice Network, 2011).

In this paper, we compare the relative impacts of legal, friend, family, and internal sanctions on minor, non-violent crimes. We refer to these transgressions as dishonest because they benefit the individual at society's expense. We define dishonest actions as those that violate a formal or informal social rule for personal gain; by this definition, lying, cheating, and stealing may all be considered facets of dishonesty. It is worth noting that some dishonest actions harm other

individuals rather than society at large; for example, lying to one's partner likely violates the (spoken or unspoken) relationship contract, but does not directly harm society. Typically, dishonest actions that harm the collective (e.g., underreporting income on one's taxes) are also subject to legal penalties; the present research focuses on violations of this nature.

While much of the existing research focuses on a single category of sanctions on crime, in the present study, we compare the relative impacts of legal, friend, family, and internal sanctions and consider their interactions. By drawing on a participant sample from five countries in distinct cultural regions, we examine whether the deterrent effects of legal, social, and internal sanctions are consistent across individuals from different cultural backgrounds.

What Deters Crime?

A sizeable body of research on the subject of what deters crime has focused on the effectiveness of legal sanctions. This research stems from deterrence theory, which posits that legal sanctions deter citizens from engaging in criminal activity. This theory, grounded in the rational actor approach, is based on the notion that people choose whether or not to commit a crime by weighing the potential benefits of getting away with it against the potential consequences of getting caught (Becker, 1968). Consequences are considered in terms of both severity of the punishment and the probability of being caught. Building on thinking of 18th century philosophers Beccaria (1963 [1764]) and Bentham (1988 [1789]), and revived in the 1960s, deterrence theory has generated much research and heated debate, with some researchers arguing that legal sanctions have no effect at all (e.g., Fattah, 1983).

Recently, Rupp (2008) conducted an impressive meta-analysis synthesizing the findings from 700 studies testing the deterrence hypothesis, spanning economics, sociology, psychology, and criminology. Detailed information about each study, including aspects of study design (cross-sectional, experimental, survey, etc.), participant sample, categories of sanctions measured, and information about the authors and journal were coded and analyzed. On the whole, this meta-analysis favored rejecting the null hypothesis that legal sanctions have no deterrent effect on crime. Furthermore, the probability of legal sanctions was found to have a greater deterrent effect than the severity of legal sanctions. In Rupp's analysis, there was also a clear pattern for legal sanctions to have stronger deterrent effects for minor, non-violent crimes (including tax evasion, speeding, and fraud) than for violent or more serious crimes (including hard drug dealing, sexual assault, and manslaughter). This pattern suggests a categorical difference in the factors deterring minor and more serious crimes. In the present paper, our research scope is limited to the factors influencing minor, non-violent crimes.

A chief criticism of deterrence theory has been its neglect of non-economic factors that may influence crime (Meier et al., 1984; Williams and Hawkins, 1986). Researchers from sociology and other traditions have suggested that non-economic sanctions have at least as much potential to impact criminal behavior (Wrong, 1961; Grasmick and Green, 1980; Mazar et al., 2008). One type of non-economic sanction considered is judgment by friends and family, which some have referred to as the threat of social embarrassment (Grasmick and Bursik, 1990; Cochran et al., 1999). Research from psychology and sociology suggests that people are highly sensitive to social evaluation (Dickerson et al., 2008). However, according to Rupp's meta-analysis, of the 2534 variables examined in survey studies, only 6.2% assessed the perceived probability of punishment by friends or family, 4.1% assessed the perceived severity of punishment by friends or family, and 2.8% assessed the perceived probability of detection by friends, family or others. Results from the meta-analysis indicated that the probability of punishment by friends or family was at least as strong a deterrent as the probability of legal punishment, and the severity of punishment by friends or family, though less powerful than the probability effects, was at least as strong a deterrent as the severity of legal punishment.

Finally, there appears to be increasing awareness that in addition to external sanctions, internal sanctions such as feelings of guilt may be important deterrents of crime. Though focused on dishonest rule violations rather than illegal actions *per se*, Mazar et al. (2008) posited that dishonesty is regulated largely by the internal desire to maintain a positive self-concept, which is weighted against the potential material benefits of breaking the rules. In support of this theory, experiments showed that increasing the flexibility with which people can categorize their dishonest actions (e.g., cheating for tokens with monetary value rather than money itself) encourages dishonesty, and conversely, that drawing attention to moral standards mitigates dishonesty. Furthermore, several experimental studies have found that increasing financial incentives for behaving dishonestly has surprisingly little impact on dishonest behavior (Wiltermuth, 2011; Gino et al., 2012; John et al., 2014; Weisel and Shalvi, 2015). For example, John et al. (2014) found that participants were just as likely to cheat on a trivia game when they were paid 5 cents per self-reported correct answer as when they were paid 25 cents per self-reported correct answer.

Considering Interactions Between Sanctions

An additional question sometimes raised by researchers is whether the deterrent effects of legal, social, and internal sanctions are independent of one another. Some scholars have raised the interesting hypothesis that the deterrent effects of legal sanctions should be most evident when moral commitments (i.e., internal sanctions) are weak (Zimring, 1971; Silberman, 1976). Evidence supporting this interaction hypothesis was reported by Silberman (1976), and more recently by Wenzel (2004), who found that in a sample of Australian citizens, penalties for tax evasion had a deterrent effect only when internal sanctions were lax. However, Grasmick and Green (1980, 1981) argued against this interaction hypothesis in favor of additive effects.

An Integrated Deterrence Framework

While many researchers who have explored the impacts of social and internal sanctions on crime have contrasted their approaches with deterrence theory, Grasmick and Bursik (1990) proposed that the deterrence framework could be extended to incorporate social and internal sanctions. They

designed a survey with questions assessing the perceived probability and severity of legal, social, and internal sanctions. Sanction threat variables, computed as the product of perceived probability and severity, were entered as predictors in regression models for three illegal actions: tax evasion, theft and drunk driving. Across the three actions, both legal sanctions and internal sanctions were significant deterrents, but internal sanctions had the stronger deterrent effect. Surprisingly, the deterrent effect of social sanctions was not significant.

Are People Deterred From Crime the Same Way Everywhere?

The limited number of studies employing Grasmick and Bursick's extended deterrence framework support their original findings that legal and internal sanctions deter crime, with internal sanctions having the stronger deterrent effect (Grasmick et al., 1993a,b; Cochran et al., 1999; Kobayashi et al., 2001). Notably, these studies have failed to provide evidence for a deterrent effect of social sanctions; the reason these effects differ from those reported in Rupp's meta-analysis is not entirely clear. Moreover, these studies have been conducted on Americans, raising the question of whether the findings are robust across cultures. (Kobayashi and colleagues' study is an exception, including both Americans and Japanese, but the researchers do not compare the strengths of deterrent effects across cultures. Wenzel (2004) also reports similar effects in an Australian sample.)

In his meta-analysis of the deterrence literature, Rupp found that the deterrent effect of legal sanctions varied according to the country under study. For example, support for the deterrence hypothesis was stronger in studies conducted in Germany and the UK than in studies conducted in Canada (Rupp, 2008). However, comparisons in Rupp's analysis were limited to select Western nations with sufficient numbers of studies testing the effects of legal deterrents. Furthermore, the deterrence effect was also found to vary according to authors' home country and country of publication, raising the possibility that the cross-country variation observed was related to author biases. Comparing culturally distinct countries within a single study overrides these issues, and allows for a more rigorous assessment of whether the relative effects of legal, social, and internal sanctions are consistent across cultures.

The Present Research

Building on the extended deterrence framework of Grasmick and Bursik (1990), we compared the deterrent effects of legal, social, and internal sanctions on minor, non-violent crimes within a single study. To compare the relative influences of these deterrents across cultures, we administered our study to an international participant sample from five countries: China, Colombia, Germany, Portugal, and USA. These countries are based in distinct cultural regions of the world, namely Confucian (China), Catholic Latin America (Colombia), Protestant Europe (Germany), Catholic Europe (Portugal), and English-speaking (USA), according to cultural mapping by Inglehart and Welzel (2010). The countries sampled differ along two broad cultural dimensions identified by Inglehart and Baker (2000) and Inglehart and Welzel (2010): traditional vs. secular-rational values and survival vs. self-expression values. Within each country, we administered a survey to two participant groups: students at public universities, and the general public at coffee shops in major cities.

We designed a survey with four sanction categories: legal, friends, family, and internal. While the threats of judgment from friends and family have traditionally been grouped together as social sanctions, we considered that judgment from friends and judgment from family might have different motivational impacts, which might vary across cultures. For example, the threat of family sanctions, but not friend sanctions, may be stronger in more traditional cultures. The first three sanction categories (legal, friends, and family) focus on negative consequences that are *external* to the individual. The final category focuses on internal consequences, namely on feelings of guilt. Other researchers used the term shame rather than guilt in referring to internal sanctions (Grasmick et al., 1993a; Kobayashi et al., 2001). In the psychological literature, guilt is typically construed as feeling badly over one's actions, while shame is typically construed as feeling badly over who one is (Tangney, 1998). Because guilt is triggered by violating internal moral standards, and may or may not induce shame, our internal sanctions measure asks about feelings of guilt rather than shame.

Participants were first asked to report the likelihood of engaging in seven minor, non-violent crimes, including parking illegally, bribing a police officer, and tax evasion. For each action, participants were asked to rate both the probability of detection and severity of punishment across each of the four sanction categories.

Our primary research questions were whether legal, social, and/or internal sanctions negatively predict the likelihood of engaging in dishonesty, and whether deterrent effects are consistent across cultures. Based on previous research suggesting the primacy of internal influences, we hypothesized that internal sanctions would have the strongest deterrent effect across cultures. In addition, we tested the interaction hypothesis that the effects of legal sanctions are stronger when internal sanctions are lax.

MATERIALS AND METHODS

This study was administered with approval from Duke University's Institutional Review Board for Non-Medical Research. All participants provided their informed written consent.

Participants

A total of 1,251 individuals completed the crime sanctions survey. To ensure that our participant sample reflected the cultures of our countries of interest, we limited our analyses to those who were native residents of each country (born in and currently residing in the country). In addition, twelve individuals were

excluded due to technical issues or internal reasons, leaving 1,100 participants in our final sample. Approximately half of the participants ($N = 586$) were students recruited from public universities, while the other half ($N = 514$) were members of the general public, recruited in coffee shops from the same cities. Participants were sampled from five countries: China, Colombia, Germany, Portugal, and USA.

Crime Sanctions Survey

All survey materials were translated into the native language of participants from each country, using a forward–backward translation procedure. Participants completed the survey individually on iPads. An instructions screen informed them that they would be asked different questions about the same actions, and that they should respond as honestly as possible. They were assured that their responses were confidential and anonymous. All participants were first asked about the likelihood that they would engage in seven minor, non-violent crimes, in the form, "How likely are you to ____?" Participants responded on continuous sliding scales ranging from 0 ("not at all likely") to 10 ("very likely").

Participants indicated how likely they would be to engage in the following actions:

(1) Omit information on your tax filings in order to pay less income tax
(2) Speed by 15% over the speed limit while driving
(3) Run a red light when nobody is around
(4) Park your car in a no parking zone
(5) Bribe a police officer to avoid getting a speeding ticket
(6) Apply for a government tax credit knowing you are not eligible for it
(7) Fake a signature of a doctor on a government document in order to get an expensive medication for free.

These questions were presented on the same screen in randomized order.

Next, participants were asked to report their perceptions of legal, social, and internal sanctions for each of the seven actions. Participants were asked about two categories of social sanctions, friends and family, resulting in four sanction categories. For each category, participants were asked about the perceived probability of being penalized for engaging in the actions with the following questions:

Legal probability: How likely would you be to get caught by the government authorities or police if you. . .

Social probability (friends): How likely would your friends be to find out if you. . .

Social probability (family): How likely would your family be to find out if you. . .

Internal probability: How likely would you be to feel guilty if you. . .

Continuous sliding scales ranged from 0 ("extremely unlikely" to 10 "extremely likely"). Furthermore, participants were asked to rate the expected severity of the legal, social, and internal consequences, as follows:

Legal severity: How bad would the legal penalty be if you. . .

Social severity (friends): How badly would your friends judge you if you. . .

Social severity (family): How badly would your family judge you if you. . .

Internal severity: How badly would you feel if you. . .

Continuous sliding scales ranged from 0 ("not bad(ly) at all") to 10 ("extremely bad(ly)").

The eight question categories were presented in random order, with the seven individual actions presented in random order within each block. In total, participants responded to 56 specific questions about legal, social and internal sanctions.

Procedure

Students at universities were recruited with flyers and posters advertising a decision-making study where they could earn between $4 and $10. At universities, the study was run in a testing room with 5–8 separate stations for participants. In coffee shops, participants were approached individually by an experimenter, who asked whether they would be interested in participating in a decision-making study with the opportunity to earn between $4 and $10. Coffee shop patrons who agreed to participate completed the survey individually from where they were seated.

Participants first completed a behavioral task on iPads, which involved rolling a virtual die twenty times (adapted from Jiang, 2013; see Mann et al., under review for further detail). Before each roll, participants were instructed to select a side of the die, either top or bottom. They were instructed to remember their chosen side, but were not asked to report choosing top or bottom until they had viewed the outcome of the roll (the screen displayed the number of dots on both top and bottom of the die). Participants were paid the equivalent of ten cents in USD for every dot on the chosen side (the amount and currency were adjusted for each country using the Purchasing Power Parity Index). Therefore, if a participant mentally selected "top" before rolling the die, and the outcome displayed one dot on the top side and six dots on the bottom side, the participant would face a choice as to whether to honestly report having chosen "top," or whether to dishonestly report having chosen "bottom". Once the participant indicated their choice, the earnings for that roll were automatically added to their total earnings, displayed at the top of the screen. With this paradigm it is impossible to know for certain whether cheating occurred on any given roll or for any given person. However, in large samples, if cheating did occur, choosing the favorable earnings side (i.e., the side with more dots) on a greater proportion of trials should be correlated with dishonesty.

When participants completed the die task, the experimenter returned and set up the crime sanctions survey on the iPad. This experimenter, who spoke participants' native language, set up the survey and instructed them to raise their hands should they have any questions. Participants indicated their responses to each question by moving bars along slider scales with their fingers. At the end of the survey, they raised their hand to indicate that they had finished. The experimenter then thanked them for participating and directed them to a payments table (for students) or paid them directly (for general public).

RESULTS

Correlations Between Likelihood of Engaging in Crime (Self-Reported) and Observed Dishonest Behavior

We first examined whether self-reported crime was related to dishonesty on the behavioral die task, in which participants could earn more money by cheating. Detailed behavioral results from the die task are reported in Mann et al. (under review); in the present paper, we present only the correlations between our behavioral measure of dishonesty and our self-report data from the crime sanctions survey. Our behavioral measure of dishonesty was the proportion of trials on the die task in which participants reported choosing the side of the die with favorable earnings. Overall, this proportion ranged from 0.56 (Portugal) to 0.60 (USA) indicating a limited but significant level of cheating in every country. We conducted a Pearson correlation between this outcome and self-reported likelihood of engaging in crime, averaged across the seven illegal actions. Across the full sample, this analysis revealed a modest but significant positive correlation ($r = 0.08$, $p = 0.012$).

Examining the correlations for each country separately revealed positive and significant correlation coefficients for Germany ($r = 0.20$, $p = 0.004$) and the USA ($r = 0.26$, $p < 0.001$), while the correlation coefficients for China, Colombia, and Portugal were not significant. Further examination indicated that these results were driven by the student samples in Germany and the USA.

Comparing Likelihood of Engaging in Crime Across Countries and Cohorts

The remaining analyses focus on our self-report data from the crime sanctions survey. We next examined whether likelihood of engaging in crime differed across countries, and across subject groups (students vs. public) within countries. **Table 1** presents the results of separate 5(Country) × 2(Cohort: student vs. public) between-subject ANOVAs conducted on each of the seven scenarios. For every scenario, reported likelihood of engaging in crime differed between countries, and results were significant at a Bonferroni-corrected probability threshold of $p = 0.007$. On the other hand, differences between cohorts were significant for only two scenarios, running a red light and falsely applying for a government tax credit, at a liberal threshold of $p = 0.05$, and for only the latter scenario at the Bonferroni-corrected threshold. Finally, the Country-by-Cohort interaction term was significant for two scenarios (speeding by 15% over the limit and running a red light), but these did not survive the Bonferroni-corrected significance threshold. Based on the limited differences observed between cohorts, along with non-significant effects for cohort in regression analyses, we combine student and public cohorts together in the analyses reported from here on. **Figure 1** shows the reported likelihood of engaging in crime for each scenario across the five countries, illustrating cultural differences.

TABLE 1 | Summary of univariate ANOVAs comparing responses across countries and cohorts regarding the likelihood of engaging in seven dishonest actions.

	Statistic	Country	Cohort	Country* Cohort
Omit information on your tax filings in order to pay less income tax	F	19.601***	0.013	0.655
	η_p^2	0.068	0.000	0.002
Speed by 15% over the speed limit while driving	F	44.898***	0.772	2.374*
	η_p^2	0.145	0.001	0.009
Run a red light when nobody is around	F	13.748***	6.633*	1.818*
	η_p^2	0.049	0.006	0.007
Park your car in a no parking zone	F	39.500***	0.706	0.910
	η_p^2	0.129	0.001	0.003
Bribe a police officer to avoid getting a speeding ticket	F	43.289***	2.551	0.529
	η_p^2	0.139	0.002	0.002
Apply for a government tax credit knowing you are not eligible for it	F	10.879***	17.983***	0.982
	η_p^2	0.039	0.017	0.004
Fake a signature of a doctor on a government document in order to get an expensive medication for free	F	12.130***	3.547	0.372
	η_p^2	0.043	0.003	0.001

All country differences were significant at a Bonferroni-corrected threshold of $p = 0.007$. At this threshold, no cohort differences were significant except applying for a tax credit knowing you are not eligible, nor were any country by cohort interactions.

*$p \leq 0.05$, **$p < 0.01$, ***$p < 0.001$.*

Deterrent Effects of Legal, Social, and Internal Sanctions

We computed legal, friend, family, and internal sanction variables by multiplying the probability and severity ratings for each action in each of the four categories. Sanction threats are commonly understood as the interaction between probability and severity of sanctions (Rupp, 2008), which derives from classical utility theory. We qualify this approach by noting that although our variables are continuous, they are not interval or ratio variables. The probability variables do not represent absolute probability scales, but rather, participants' perceptions of probability. Acknowledging the limitations of multiplying ordinal variables, for ease of interpretation and for comparison with existing theory and research, we followed tradition in multiplying self-reported probability and severity values to compute the sanction threat variables (e.g., Grasmick and Bursik, 1990; Kobayashi et al., 2001; Wenzel, 2004). For the remaining analyses, we structured our data such that each row represented a particular subject's response to a particular question.

We first examined the relative importance of the four types of sanctions across all subjects by running linear mixed effects analyses with data from all subjects and questions. These analyses were run in R Core Team (2014), using the lme4 package (Bates et al., 2014). *P*-values were computed with the Satterthwaite approximation, using the lmerTest package (Kuznetsova et al.,

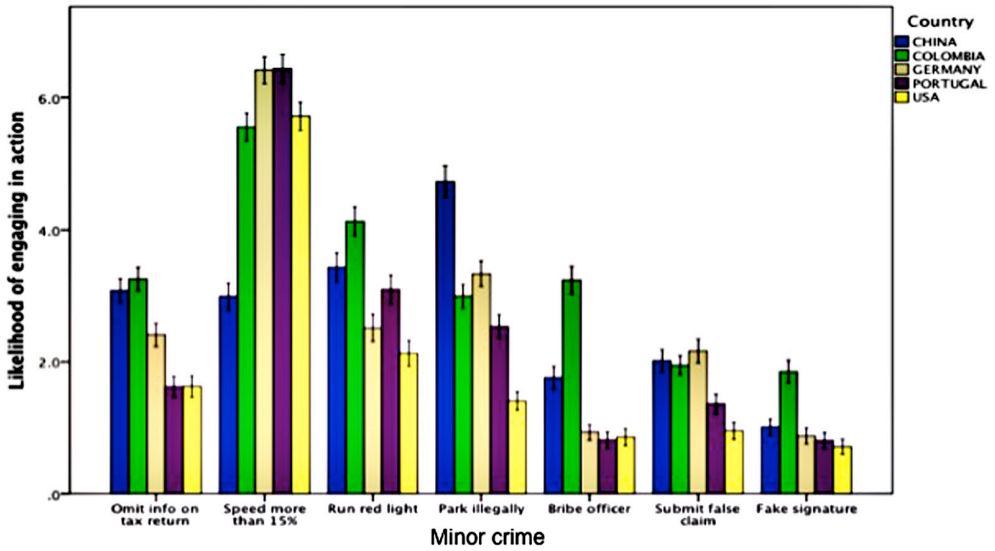

FIGURE 1 | Self-reported likelihood of engaging in seven specific dishonest actions across countries. Error bars represent ±1 SEM.

2014). Models were estimated using a maximum likelihood (ML) approach. To facilitate interpretation of the parameter estimates, all fixed effects variables and the dependent measure were first standardized to have a mean of 0 and standard deviation of 1.

Results from three mixed effect models are reported in **Table 2**. As a baseline, we ran an initial model with demographic variables (gender, age, minority status, relative earnings, religiosity, and mistrust of others) entered as fixed effects, and likelihood of engaging in crime entered as the dependent measure (Model 1). To account for non-independent responses, item, country, and subjects nested within country were entered as random effects variables. This analysis showed significant effects for gender, age, relative earnings, and mistrust in others. Women were less likely to engage in crime than men, although this did not hold up in subsequent models. Older individuals reported being less likely to engage in crime, while those with higher relative earnings reported being more likely to engage in crime overall. This finding aligns with the work by Piff et al. (2012), which suggests that upper class individuals are less ethical than lower class individuals (See also Ariely and Mann, 2013; Trautmann et al., 2013). Finally, as others have found (Uslaner and Badescu, 2004; Neville, 2012), mistrust in others was related to greater likelihood of engaging in crime.

Model 2 built on Model 1 to examine the effects of external and internal sanction threats. Including legal, friends, family, and internal sanctions as continuous fixed effect variables resulted in a highly significant model improvement over Model 1, according to a log likelihood ratio test ($\chi^2_{(4)} = 2231.6$, $p < 0.001$). As can be seen from the table, beta values for legal and internal sanctions were negative and highly significant, indicating that the greater the sanction threat, the lower an individual's reported likelihood of engaging in crime. Although both legal and internal sanctions predicted unique variance in the model, it is also worth noting that the beta value for internal sanctions ($b = -0.398$; $t(5488) = -27.253$) was five times the magnitude of the beta value for legal sanctions ($b = -0.091$, $t(5599) = -6.575$). In contrast, beta values for friends and family sanctions, though modest and only marginally significant, were positive in sign, indicating that greater threats of social judgment, whether from friends or family, predicted *greater* likelihood of engaging in crime. We return to this finding in the Discussion section.

Finally, Model 3 built on Model 2 by including two-way interaction terms for the sanction threats as fixed effect variables (interaction terms were computed from the standardized sanction threat variables). Including interaction terms led to significant model improvement over Model 2 ($\chi^2_{(6)} = 106.7$, $p < 0.001$). We were interested in testing the interaction hypothesis that when internal sanctions (i.e., feelings of guilt) are weak, legal sanctions have a stronger deterrent effect on crime. In support of this hypothesis, we observed a significant, positive interaction between internal sanctions, and legal sanctions. Similar findings were reported by Silberman (1976), and Wenzel (2004). Grasmick and Green (1980) also reported results that were similar in direction though not significant.

To further explore this effect, we conducted follow-up moderation analyses for each of the seven crimes, using Hayes' process model which follows Baron and Kenny's (1986) approach (Hayes, 2013). Internal sanctions moderated the effect of legal sanctions for four of the seven crimes (speeding, running a red light, parking illegally, and bribing an officer). For each of these crimes, the negative effect of legal sanctions was stronger when internal sanctions were weak.

In Model 3, the effect of friend sanction threats was positive and significant, and the effect of family sanction threats positive though not significant. In order to gain insight into the unexpected positive relationship between social sanction threats and likelihood of illegal actions, we conducted an additional linear mixed model analysis with standardized probability and

TABLE 2 | Results from linear mixed effects models with ML estimation for likelihood of engaging in crime.

	Model 1		Model 2		Model 3	
Fixed effects	***b***	***p***	***b***	***p***	***b***	***p***
(Intercept)	0.005	0.979	0.005	0.000***	−0.042	0.000***
FEMALE	−0.037	0.029*	0.010	0.525	0.011	0.495
Age	−0.084	0.000***	−0.036	0.027*	−0.037	0.024*
MINORITY	−0.012	0.495	−0.017	0.283	−0.020	0.223
Relative Earnings	0.057	0.001***	0.053	0.001***	0.051	0.001**
Religiosity	−0.006	0.728	0.018	0.285	0.017	0.311
Mistrust	0.043	0.012*	0.033	0.037*	0.031	0.057†
LEGAL			−0.091	0.000***	−0.122	0.000***
FRIEND			0.026	0.056†	0.037	0.029*
FAMILY			0.025	0.080†	0.017	0.437
INTERNAL			−0.398	0.000***	−0.397	0.000***
LEGAL*FRIENDS					−0.024	0.071†
LEGAL*FAMILY					−0.021	0.108
LEGAL*INTERNAL					0.113	0.000***
FRIEND*FAMILY					−0.009	0.408
FRIEND*INTERNAL					0.007	0.657
FAMILY*INTERNAL					0.023	0.106
Random effects	**σ**		**σ**		**σ**	
Subject*Country	0.400		0.371		0.377	
Item	0.460		0.301		0.292	
Country	0.143		0.126		0.133	
Residual	0.776		0.709		0.700	
Log-likelihood	−7836		−6721		−6667	
Likelihood ratio test against previous model			$\chi^2_{(4)}$ =2231.6	0.000***	$\chi^2_{(6)}$ =106.76	0.000***

All models include subject, item, and country as random effects variables, with subject nested within country. Fixed effect variables and the outcome variable were standardized for ease of interpretation. Model 1 includes demographic variables of interest as fixed effect terms. Model 2 additionally includes the four sanction variables, resulting in a highly significant model improvement. Model 3 includes two-way interactions terms for the sanction variables, again resulting in highly significant model improvement. From Models 2 and 3, both internal sanctions and legal sanctions show significant deterrent effects on crime, though the effect of internal sanctions is approximately four times greater. Friend and family sanctions are positively related to crime (significantly so for friend sanctions). A highly significant positive interaction between legal and internal sanctions indicates that the deterrent effect of legal sanctions is stronger when internal sanctions are low.
†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

severity sanction variables entered as separate fixed effect variables (**Table 3**). Demographic variables were also included in the model, with item, country, and subjects nested within country again entered as random effects variables. This analysis revealed that the probability variables for both friends and family sanctions, where subjects rated how likely their friends or family would be to find out if they acted illegally, were significant *positive* predictors of illegal actions. The severity of family judgment was a significant deterrent of illegal actions, while the severity of friends' judgment did not significantly predict illegal action.

Rupp's meta-analysis and common consensus indicate that the probability of legal sanctions has a stronger deterrent effect than the severity of legal sanctions. In contrast, in our data, sanction severity had a stronger deterrent effect than sanction probability, for both the legal and internal sanction categories.

Do the Effects of Sanctions Vary Across Countries?

Our next question was whether the deterrent effects of legal, friend, family, and internal sanctions were consistent or variable across countries. **Table 4** presents the results of linear mixed models conducted separately for each country. Standardized demographics, sanction variables, and two-way sanction interaction terms were entered as fixed effect predictors, with subject and item entered as random effects. Notably, the effect of relative earnings on engaging in crime was significant only for China and Colombia, whereas for the American sample, the effect of relative earnings was negative and non-significant. Thus, when examined at the country level, our data diverges from Piff et al. (2012) finding that upper class individuals demonstrated more unethical behavior than lower class individuals in an American sample.

As can be seen from the table, the deterrent effect of internal sanctions was highly significant across all five countries. The deterrent effect of legal sanctions was significant in China, Germany, and USA, marginally significant in Portugal, and not significant in Colombia. Finally, the positive interaction between legal and internal sanctions was significant in every country except Colombia.

TABLE 3 | Results from a linear mixed effects models (ML estimation) for likelihood of engaging in crime, with probability and severity ratings for legal, friend, family, and internal sanctions entered as predictors, in addition to demographic variables.

Fixed effects	***b***	***p***
(Intercept)	0.006	0.962
FEMALE	0.014	0.387
Age	−0.020	0.223
MINORITY	−0.017	0.296
Relative Earnings	0.039	0.012*
Religiosity	0.027	0.092†
Mistrust	0.029	0.063†
Legal (Probability)	−0.038	0.002**
Legal (Severity)	−0.071	0.000***
Friend (Probability)	0.091	0.000***
Friend (Severity)	0.010	0.527
Family (Probability)	0.093	0.000***
Family (Severity)	−0.114	0.000***
Internal (Probability)	−0.123	0.000***
Internal (Severity)	−0.280	0.000***
Random effects	**σ**	
Subject*Country	0.372	
Item	0.262	
Country	0.141	
Residual	0.676	
Log likelihood	−6474	

Fixed effects variables and the outcome variable were standardized for ease of interpretation. Subject, item, and country were as random effects variables, with subject nested within country. For legal and internal sanctions, both probability and severity ratings were negatively related to crime, with severity ratings having somewhat stronger effects. For friend and family sanctions, probability of being detected was positively related to crime; severity of judgment from family was negatively related to crime, while severity of judgment from friends was not significant.

*†p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.*

To determine whether the strength of sanction threats varied significantly across countries, we ran a linear mixed effects model with standardized sanction variables and individual countries entered as fixed effect variables, in addition to sanction threat by country interaction terms. Country variables were coded using effect coding instead of dummy coding such that each country's mean could be compared against the grand mean. As is the case for dummy coding, with effect coding for k groups, only k–1 groups can be estimated according to the degrees of freedom. In order to report parameter estimates for all five countries, we ran the linear mixed effects model twice with a different country excluded from estimation each time, and reported the parameters for all five countries in **Table 5**. Other parameters in the model are not affected by the country that is excluded from effect coding.

As can be seen from **Table 5**, country main effects were significant only for Colombia and USA; overall, Colombians reported greater-than-average likelihood of engaging in illegal actions ($b = 0.614$, $p < 0.001$), while Americans reported less-than-average likelihood. Sanction by country interactions terms allowed us to address the question of whether the deterrent effects of sanctions varied according to country. Legal sanctions were found to have stronger deterrent effects for China and weaker deterrent effects in Colombia. The reverse deterrent effect of friend sanctions was particularly strong in China relative to the other countries (positive interaction term) whereas a negative interaction term was observed for USA. With regard to family sanctions, no significant differences were observed across countries. Finally, internal sanctions had significantly stronger deterrent effects in Germany, and marginally stronger deterrent effects in Colombia, whereas in the USA, internal sanctions were weaker relative to other countries.

Do Deterrent Effects Vary Across Actions?

Until this point, variation in specific crimes was treated as a nuisance variable. To compare the deterrent effects of sanction threats across the seven actions, we conducted separate linear regression analyses for each action. Legal, friend, family, and internal sanctions for the specific crime were entered as predictors, along with demographic variables (predictor variables were unstandardized, as the standardized beta values are computed for these models). First, the series of linear regression analyses was run on the full sample, not distinguishing subjects based on country. These analyses were then repeated on subjects from each of the five countries separately.

The beta values for legal, friend, family and internal sanction threats for each series of regression analyses are depicted in **Figure 2**. As can be seen from the figure, with limited exceptions, the deterrent effects of internal sanction threats are non-overlapping with the deterrent effects of the other categories of sanction threats.

DISCUSSION

Building on a substantial literature examining the deterrence hypothesis, the present research compared the effectiveness of legal, social (both friend and family), and internal sanctions on deterring minor, non-violent crimes in an international sample spanning five countries. Replicating the findings of others (Grasmick and Bursik, 1990; Grasmick et al., 1993a,b; Cochran et al., 1999; Kobayashi et al., 2001; Wenzel, 2004), we found internal sanctions to have the strongest deterrent effect on crime. This pattern was observed in every country studied, indicating that the primacy of internal sanctions is robust across cultures. In line with deterrence research, legal sanctions were also found to have a significant though weaker overall effect. The effect of legal sanctions was significant in China, Germany, and USA, marginally significant in Portugal, and non-significant in Colombia, suggesting variability across cultures in the extent to which legal sanctions effectively deter crime. The relative effects of internal and legal deterrents were also robust across actions, with internal sanctions usurping legal sanctions for every action in every country, with only one exception (bribing a police officer by Americans).

Some researchers have proposed that the deterrent effects of legal sanctions are stronger when internal sanctions are lax,

TABLE 4 | Results from linear mixed effects models (ML estimation) for likelihood of engaging in crime, conducted separately for each country.

	China		Colombia		Germany		Portugal		USA	
Fixed effects	***b***	***p***	***b***	***p***	***b***	***p***	***b***	***p***	***b***	***p***
(Intercept)	0.009	0.921	0.147	0.246	−0.116	0.406	−0.086	0.624	−0.254	0.151
FEMALE	−0.031	0.504	−0.036	0.356	−0.010	0.745	0.023	0.515	0.071	0.025*
Age	−0.173	0.054†	−0.067	0.108	−0.036	0.253	−0.072	0.053†	−0.013	0.595
MINORITY	−0.111	0.029*	−0.043	0.354	−0.012	0.818	−0.053	0.212	0.023	0.273
Relative Earnings	0.192	0.000***	0.111	0.005**	0.033	0.234	0.042	0.241	−0.036	0.241
Religiosity	−0.033	0.509	0.075	0.044*	0.024	0.463	0.043	0.223	−0.022	0.483
Mistrust	0.017	0.690	0.044	0.207	−0.004	0.909	0.076	0.040*	0.037	0.266
LEGAL	−0.235	0.000***	−0.019	0.554	−0.116	0.000***	−0.055	0.053†	−0.127	0.000***
FRIEND	0.102	0.006**	0.027	0.422	0.025	0.527	−0.004	0.910	−0.067	0.083†
FAMILY	0.077	0.099†	0.036	0.269	−0.058	0.083†	0.067	0.050	0.033	0.275
INTERNAL	−0.418	0.000***	−0.457	0.000***	−0.433	0.000***	−0.373	0.000***	−0.244	0.000***
LEGAL*FRIEND	−0.024	0.461	−0.045	0.109	−0.011	0.720	−0.015	0.587	0.005	0.856
LEGAL*FAMILY	−0.040	0.275	0.006	0.831	−0.023	0.439	−0.028	0.278	−0.068	0.007**
LEGAL*INTERNAL	0.179	0.000***	0.021	0.456	0.143	0.000***	0.090	0.000***	0.085	0.000***
FRIEND*FAMILY	−0.002	0.940	−0.021	0.400	−0.008	0.755	0.000	0.984	0.008	0.674
FRIEND*INTERNAL	−0.019	0.605	0.028	0.365	0.005	0.897	0.013	0.679	0.018	0.558
FAMILY*INTERNAL	−0.005	0.893	−0.014	0.656	0.085	0.006**	−0.018	0.541	0.039	0.137
Random effects	**σ**		**σ**		**σ**		**σ**		**σ**	
Subject	0.175		0.404		0.283		0.335		0.358	
Item	0.028		0.290		0.334		0.432		0.408	
Residual	0.509		0.730		0.654		0.652		0.595	

The outcome variable was standardized, and standardized demographics, legal, friend, family, and internal sanctions, and two-way sanction interaction terms were entered as fixed effect variables. Subject and item were entered as random effect variables.
*†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

though others have argued in favor of purely additive effects (Grasmick and Green, 1980, 1981). Supporting the interaction hypothesis, we observed a significant positive interaction between legal and internal sanction threats, an effect also observed by Wenzel (2004) in his study of tax evasion among Australian citizens. In our international sample, the interaction was evident in every country except Colombia. Follow-up moderation analyses showed that the effect of legal sanctions was significant only when internal sanctions were lax; however, the moderation was significant for only four of the seven illegal actions (in contrast to Wenzel's findings, the effect was not significant for tax evasion). These results suggest that the interaction between legal and internal sanctions may depend on the particular action.

Social Influences on Crime

An unexpected finding was the positive relationship observed between social sanctions and crime. Overall, the effect of friend sanctions was positive and significant. Examining countries separately, a significant or marginally significant positive effect for either friend or family sanctions was observed in every country except Colombia. To better understand these effects, we conducted additional analyses with probability and severity sanction variables entered as separate predictors. In every country, probability of being found out by friends was positively related to likelihood of acting illegally; the same was true for probability of being found out by family in every country except Germany. Although this result was not anticipated, we speculate that both probability of engaging in crime and probability of being found out by friends and family may be related to a third underlying variable, namely the extent to which the action is normative. For example, if bribing a police officer is a widely practiced behavior in a particular society, an individual in that society may be more likely to practice the behavior, and her friends may be more likely to know about it, than an individual in a society where bribing police is not normative. In line with the hypothesis that social norms influence dishonesty, Gino et al. (2009) found that individuals were more likely to cheat on a test after observing an in-group member cheat, while observing an out-group member cheat had the opposite influence on dishonest behavior.

Another possibility is that people who engage in crime give more thought to the possibility of others finding out about their actions. For example, if a person regularly parks illegally, she may be more likely to think about (and overestimate) the possibility of being found out by friends relative to others who have rarely contemplated this crime. Thus, normativity and degree of cognitive reflection are two potential explanations for the observed positive relationship between probability of being found out and probability of engaging in crime. Since we cannot test third variable explanations with the given data, we recommend that future research examining the relationship between social sanctions and dishonest behavior incorporate these variables.

TABLE 5 | Results from a linear mixed effects model (ML estimation) for likelihood of engaging in crime, with demographics, sanction variables, and countries included as fixed effect variables.

Fixed effects	*b*	*p*
(Intercept)	2.578	0.000***
FEMALE	0.033	0.499
Age	−0.120	0.016*
MINORITY	−0.047	0.339
Relative Earnings	0.153	0.002**
Religiosity	0.051	0.310
Mistrust	0.106	0.029*
LEGAL	−0.329	0.000***
FRIEND	0.082	0.048*
FAMILY	0.073	0.105
INTERNAL	−1.201	0.000***
CHINA	0.120	0.282
COLOMBIA	0.614	0.000***
GERMANY	−0.071	0.461
PORTUGAL	−0.129	0.185
USA	−0.534	0.000***
LEGAL*CHINA	−0.497	0.000***
LEGAL*COLOMBIA	0.299	0.000***
LEGAL*GERMANY	0.004	0.960
LEGAL* PORTUGAL	0.127	0.107
LEGAL*USA	0.067	0.417
FRIEND*CHINA	0.296	0.000***
FRIEND*COLOMBIA	−0.015	0.851
FRIEND*GERMANY	−0.035	0.706
FRIEND*PORTUGAL	−0.049	0.560
FRIEND*USA	−0.198	0.012*
FAMILY*CHINA	0.027	0.791
FAMILY*COLOMBIA	0.014	0.859
FAMILY*GERMANY	−0.111	0.216
FAMILY*PORTUGAL	0.026	0.760
FAMILY*USA	0.043	0.604
INTERNAL*CHINA	0.083	0.414
INTERNAL*COLOMBIA	−0.153	0.055†
INTERNAL*GERMANY	−0.213	0.007**
INTERNAL*PORTUGAL	−0.131	0.107
INTERNAL*USA	0.414	0.000***
Random effects	**σ**	
Subject	1.123	
Item	0.915	
Residual	2.148	
Log likelihood	−13158	

Demographic variables, sanction variables, and the outcome variable were standardized for ease of interpretation. Two-way interaction terms between sanction and country variables were also included in the model. Subject and item were entered as random effect variables. Effects coding was used for countries, such that the reported parameter estimates compare each country's mean against the grand mean. The analysis was run twice with a different country excluded in the deviation time each time, so that parameter estimates for all five countries could be reported.

†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

We are not aware of any other study reporting a positive relationship between social sanction threats and likelihood of engaging in crime. Grasmick and Scott (1982) observed deterrent effect of social sanctions on crime, while several other studies comparing legal, social, and internal sanctions have failed to find deterrent effects of social sanctions (Grasmick and Bursik, 1990; Grasmick et al., 1993a; Cochran et al., 1999; Kobayashi et al., 2001). Taken together, what can we make of these results? Do they imply that threat of social judgment does not impact likelihood of engaging in crime? Such a conclusion seems highly unlikely in light of a vast body of research illustrating the power of social norms (Cialdini and Goldstein, 2004; Fehr and Fischbacher, 2004). We propose instead that the power of social norms occurs primarily through their internalization as moral standards by members of society (Campbell, 1964). When individuals identify with their society, they adopt society's moral standards as personal moral standards (Wenzel, 2004). The threat of personal judgment (feelings of guilt) for one's own transgressions then becomes a more effective deterrent than the judgment of friends or family. In support of this theory, Wenzel (2004) found in an Australian sample that social norms had a significant effect on tax evasion only for those who did not identify as Australian (i.e., those who presumably did not internalize the prevailing standards).

Some scholars have proposed that legal sanctions deter crime not through material disincentives but by increasing the level of social condemnation that results from a dishonest action (Tittle and Logan, 1973; Williams and Hawkins, 1986). According to this theory, if a person acts dishonestly, other people will judge her more harshly for her action if it is against the law, and it is this increased threat of social judgment that accounts for the legal deterrent effect. Interestingly, we observed a marginally significant negative interaction between legal and social sanctions, implying that the legal deterrents were more effective when social sanctions were stronger. In his study of tax evasion, Wenzel (2004) observed a similar effect (though it was only evident for those who did not identify as Australian citizens). These results provide tentative evidence for synergistic effects when legal and social sanctions operate in tandem.

Implications

Kobayashi et al. (2001) examined differences in workplace compliance between Japanese and American employees, and found that these differences could be accounted for by differences in perceived internal, social and management (regulatory) sanctions. In contrast, while we observed country differences in terms of likelihood of engaging in specific crimes, these cross-cultural differences in likelihood of engaging in crime were not entirely accounted for by differences in sanctions. For six of the seven actions in our study, differences in legal sanctions across countries explained some of the variation in country-level differences in crime, while differences in social and internal sanctions were unrelated to country variation. These results raise the interesting possibility that cultural drivers of dishonesty are not entirely captured by sanctions. For example, it is possible that cultural differences in internal or external reward associated with dishonesty account for variation in crime. Further research is needed to understand whether cross-cultural variation in crime

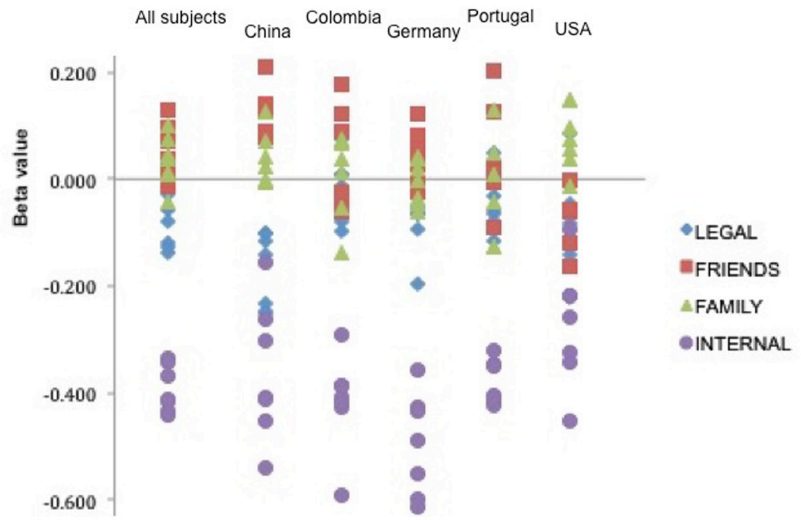

FIGURE 2 | Summary of the beta values for legal, friends, family, and internal product variables entered as predictor variables in linear regression analyses. For each item, sanction product variables were entered as predictors with self-reported likelihood of engaging in the action entered as the dependent measure.

is best accounted for by differences in sanctions or differences in other variables.

From a policy perspective, our findings raise the important question of whether policy efforts can change people's internal moral commitments to honesty and socially upright behavior. In a longitudinal study on drunk driving, Grasmick et al. (1993a) measured intentions to engage in drunk driving in 1982 and 1990, along with perceived legal, social, and internal sanctions, among residents of Oklahoma City. This 8-year interval was characterized by social efforts aimed at reducing drunk driving (for example, Mothers Against Drunk Driving rose to prominence during this time), as well as harsher legal sentences (Jacobs, 1989; Ross, 1994). The study found that intentions to engage in drunk driving indeed diminished over the 8-year period – but that the reduction was primarily accounted for by the threat of internal sanctions, rather than perceived threats of social or legal sanctions. These results suggest that over time, efforts at changing policy and/or social attitudes may translate into internalized morals.

Limitations and Future Directions

Our study should be qualified in light of limitations. Our data were collected using survey methodology. Directly asking participants to assess the probability and severity of sanction threats after reporting the perceived likelihood of engaging in minor, non-violent crimes has the advantage of enabling direct comparison of legal, social, and internal sanctions. However, this methodology yields results that are correlational, and based on self-report. It is natural to wonder whether social desirability biases influence reports of dishonest or illegal behavior. Although we cannot rule out this possibility, we note that self-report methodologies are commonly used to assess dishonesty (Grasmick et al., 1993a; DePaulo et al., 1996; Cochran et al., 1999; Kobayashi et al., 2001; Ennis et al., 2008). We do acknowledge the possibility that social desirability bias may vary by country (Bernardi, 2006). However, social desirability bias should not affect our main results provided that it does not differentially impact reports of probability or severity of legal, social, or internal sanction threats.

Future research may provide complementary evidence by comparing the strengths and interplay of legal, social, and internal sanctions using experimental methods. For example, future research might examine the hypothesis that internal sanctions derive from social norms by manipulating whether a particular dishonest action is condemned by in-group or out-group members, and then measuring participants' (a) likelihood of engaging in the dishonest action themselves, and (b) judgment of others who engage in the dishonest action. Furthermore, researchers may vary the extent to which social condemnation of an action is seen as universal or variable, and then assess participants' own views of the action. Finally, it would be interesting to test the interaction between legal and internal sanctions experimentally. For example, researchers might examine whether manipulating the perceived probability and severity of legal sanctions for illegal downloading differentially impacts downloading behavior for participants with strong versus weak personal morals against piracy.

In addition, our data highlight the need for further research into how income and social class impact moral behavior. There has been some discussion in the literature concerning whether social status influences unethical behavior. This discussion was spurred by Piff and colleagues' findings that upper class individuals were more likely to violate the law than lower class individuals, and that being primed with an upper class mindset encourages greater levels of unethical behavior (Piff et al., 2012). These findings were based on data from American samples. In the present study, we observed an overall positive relationship between relative earnings and likelihood of engaging

in minor, non-violent crimes. However, when examining our data at the country level, the relationship was significant and positive for China and Colombia only, whereas the correlation for Americans was negative and non-significant. In another study by Trautmann et al. (2013) employing a representative sample of Dutch participants, the authors did not observe a positive correlation between income and unethical behavior. Trautmann et al. (2013) argued that the relationship between class and unethical behavior is more complex than posited by Piff et al. (2012), a conclusion that appears to be supported by our data (see also Ariely and Mann, 2013). However, we note that both Trautmann et al. (2013) study and the present study provide correlational evidence, whereas Piff et al. (2012) have reported causal evidence in which priming an upper-class mindset leads to more unethical behavior. Further experimental research employing participant samples from different countries is needed to better understand the interesting and potentially complex relationship between income, class, and moral behavior.

CONCLUSION

Our findings suggest that across societies and cultures, internalized moral standards exert the most powerful restraints on dishonest behavior (see also Campbell, 1964). Policy efforts aimed at promoting moral internalization may be more effective than efforts aimed at increasing the frequency or probability of legal sentences. However, the process by which internalization occurs remains poorly understood, and marks an important direction for future research aimed at reducing crime and enhancing social welfare.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct, and intellectual contribution to the work, and approved it for publication. All authors were involved in data collection and theoretical framing. Data analysis was conducted by HM.

ACKNOWLEDGMENTS

Thanks to Jonathan Schulz for thoughtful discussions and Samuel Iglesias for updating the die task with each new translation. For their help translating materials, thanks to Zishu Chen, Sophie Guo, Franziska Greiner, Tobias Kuntze and, Natacha Barreto, Ana Sofia Braga, and Miguel Abrantes. Thanks to Zishu Chen, Siyang Wang, Kitty Vorisek, and Miguel Abrantes for their help in making local connections.

Thanks to Iván Hernández, Mário Augusto Boto Ferreira, Zhonghua Cai, and Yanchen Bi for their support with logistics. Thanks also to our amazing research assistants in China: Tingting An, Yanchen Bi, Zishu Chen, Yuan Gao, Qihuan Song, Xicheng Teng, Likun Wang, Junjie Yin; Colombia: Rocío del Pilar Alba, Paola García Arévalo, Laura Lugo, Laura Beatriz Roa; Germany: Rene Cyranek, Leonard Doyle, Oliver Gmuender, Lisa Kitzinger, Julia Memmert, Angelica Schmidt, Lucia Sommerer, Johanna Staffler, Max Straka; Portugal: Ana Sofia Braga, Margarida Cipriano, Pedro Garcia-Marques, Ana Lapa, Manuel Oliveira, Joana Reis, Mariana Sequeira; and USA: Omar Daouk, Sunny Kang, Amy Taggart, Gloria Tomlinson, Alev Uneri. We are grateful to Dan Ariely for his generous support of this project.

REFERENCES

Ariely, D., and Mann, H. (2013). A bird's eye view of unethical behavior: commentary on Trautmann et al. (2013). *Perspect. Psychol. Sci.* 8, 498–500. doi: 10.1177/1745691613498907

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models Using 'Eigen' and S4. R Package Version 1.1-7*. Available at: http://CRAN.R-project.org/package=lme4

Beccaria, C. (1963 [1764]). *On Crimes and Punishment*. Indianapolis, IN: Bobbs-Merrill.

Becker, G. S. (1968). Crime and punishment: an economic approach. *J. Polit. Econ.* 76, 169–217. doi: 10.1086/259394

Bentham, J. (1988 [1789]). *The Principles of Morals and Legislation*. Amherst, NY: Prometheus Books.

Bernardi, R. A. (2006). Associations between hofstede's cultural constructs and social desirability response bias. *J. Bus. Ethics* 65, 43–53. doi: 10.1007/s10551-005-5353-0

Campbell, E. Q. (1964). The internalization of moral norms. *Sociometry* 27, 391–412. doi: 10.2307/2785655

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Cochran, J. K., Florida, S., Wood, P. B., and Sellers, C. S. (1999). Shame, embarrassment, and formal sanction threats: extending the deterrence/rational choice model to academic dishonesty. *Sociol. Inq.* 69, 91–105. doi: 10.1111/j.1475-682X.1999.tb00491.x

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

Dickerson, S. S., Mycek, P. J., and Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychol.* 27, 116–121. doi: 10.1037/0278-6133.27.1.116

Ennis, E., Vrij, A., and Chance, C. (2008). Individual differences and lying in everyday life. *J. Soc. Pers. Relat.* 25, 105–118. doi: 10.1177/0265407507086808

Fattah, E. A. (1983). A critique of deterrence research with particular reference to the economic approach. *Can. J. Crimol.* 25, 79–90.

Fehr, E., and Fischbacher, U. (2004). Social norms and human cooperation. *Trends Cogn. Sci.* 8, 185–190. doi: 10.1016/j.tics.2004.02.007

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gino, F., Ayal, S., and Ariely, D. (2012). *Self-Serving Altruism? When Unethical Actions That Benefit Others Do Not Trigger Guilt*. Boston, MA: Harvard Business School, 13–28.

Grasmick, H. G., Bursick, R. J. J., and Arneklev, B. J. (1993a). Reduction in drunk driving as a response to increased threats of shame, embarrassment, and legal sanctions. *Criminology* 31, 41–67. doi: 10.1111/j.1745-9125.1993.tb01121.x

Grasmick, H. G., Blackwell, B. S., Bursik, R. J., and Mitchell, S. (1993b). Changes in perceived threats of shame, embarrassment, and legal sanctions for interpersonal violence, 1982-1992. *Violence Vict.* 8, 313–325.

Grasmick, H. G., and Bursik, R. J. (1990). Conscience, significant others, and rational choice: extending the deterrence model. *Law Soc. Rev.* 24, 837–862. doi: 10.2307/3053861

Grasmick, H. G., and Green, D. E. (1980). Legal punishment, social disapproval and internalization as inhibitors of illegal behavior. *J. Crim. Law Criminol.* 71, 325–335. doi: 10.2307/1142704

Grasmick, H. G., and Green, D. E. (1981). Deterrence and the morally committed. *Sociol. Q.* 22, 1–14. doi: 10.1111/j.1533-8525.1981.tb02204.x

Grasmick, H. G., and Scott, W. J. (1982). Tax evasion and mechanisms of social control: a comparison with grand and petty theft. *J. Econ. Psychol.* 2, 213–230. doi: 10.1016/0167-4870(82)90004-6

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York, NY: Guilford Press.

Inglehart, R., and Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *Am. Sociol. Rev.* 65, 19–51. doi: 10.2307/2657288

Inglehart, R., and Welzel, C. (2010). Changing mass priorities: the link between modernization and democracy. *Perspect. Polit.* 8, 551–567. doi: 10.1017/S1537592710001258

Jacobs, J. B. (1989). *Drunk Driving: An American Dilemma*. Chicago, IL: University of Chicago Press.

Jiang, T. (2013). Cheating in mind games: the subtlety of rules matters. *J. Econ. Behav. Organ.* 93, 328–336. doi: 10.1016/j.jebo.2013.04.003

John, L. K., Loewenstein, G., and Rick, S. I. (2014). Cheating more for less: upward social comparisons motivate the poorly compensated to cheat. *Organ. Behav. Hum. Decision Process.* 123, 101–109. doi: 10.1016/j.obhdp.2013.08.002

Kobayashi, E., Grasmick, H., and Friedrich, G. (2001). A cross-cultural study of shame, embarrassment, and management sanctions as deterrents to noncompliance with organizational rules. *Commun. Res. Rep.* 18, 105–117. doi: 10.1080/08824090109384788

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2014). *LmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models. R Package, Version 2.0-3.*

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Market. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Meier, R. F., Burkett, S. R., and Hickman, C. A. (1984). Sanctions, peers, and deviance: preliminary models of a social control process. *Sociol. Q.* 25, 67–82. doi: 10.1111/j.1533-8525.1984.tb02239.x

Neville, L. (2012). Do economic equality and generalized trust inhibit academic dishonesty? Evidence from state-level search-engine queries. *Psychol. Sci.* 23, 339–345. doi: 10.1177/0956797611435980

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., and Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4086–4091. doi: 10.1073/pnas.1118373109

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ross, H. L. (1994). *Confronting Drunk Driving: Social Policy for Saving Lives*. New Heven, CT: Yale University Press.

Rupp, T. (2008). *Meta Analysis of Crime and Deterrence: A comprehensive Review of Literature*. Ph.D.thesis, Technische Universität Darmstadn, Darmstadt.

Silberman, M. (1976). Toward a theory of criminal deterrence. *Am. Sociol. Rev.* 41, 442–461. doi: 10.2307/2094253

Tangney, J. P. (1998). "How does guilt differ from shame?," in *Guilt and Children*, ed. J. Bybee (San Diego, CA: Acadamic Press), 1–17.

The Tax Justice Network (2011). *The Cost of Tax Abuse: A Briefing Paper on the Cost of Tax Evasion Worldwide*. Available at: http://www.taxjustice.net/2014/04/01/cost-tax-abuse-2011/

Tittle, C. R., and Logan, C. H. (1973). Sanctions and deviance: evidence and remaining questions. *Law Soc. Rev.* 7, 371–392. doi: 10.2307/3052920

Trautmann, S. T., van de Kuilen, G., and Zeckhauser, R. J. (2013). Social class and (Un)ethical behavior: a framework, with evidence from a large population sample. *Perspect. Psychol. Sci.* 8, 487–497. doi: 10.1177/1745691613491272

Uslaner, E. M., and Badescu, G. (2004). "Honesty, trust, and legal norms in the transition to democracy: why Rothstein is better able to explain Sweden than Romania," in *Creating Social Trust in Post-Socialist Transition*, eds J. Kornai, B. Rothstein, and S. Rose-Ackerman (Basingstoke: Palgrave Macmillan), 31–53.

Weisel, O., and Shalvi, S. (2015). The collaborative roots of corruption. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10651–10656. doi: 10.1073/pnas.1423035112

Wenzel, M. (2004). The social side of sanctions: personal and social norms as moderators of deterrence. *Law Hum. Behav.* 28, 547–567. doi: 10.1023/B:LAHU.0000046433.57588.71

Williams, K. R., and Hawkins, R. (1986). Perceptual research on general deterrence: a critical review. *Law Soc. Rev.* 20, 545–572. doi: 10.2307/3053466

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organ. Behav. Hum. Decision Process.* 115, 157–168. doi: 10.1016/j.obhdp.2010.10.001

Wrong, D. H. (1961). The oversocialized conception of man in modern sociology. *Am. Soc. Rev.* 26, 183–193. doi: 10.2307/2089854

Zimring, F. E. (1971). *Perspectives on Deterrence*, Vol. 2. Rockville, MD: National Institute of Mental Health, Center for Studies.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 May 2016
doi: 10.3389/fpsyg.2016.00550

Reactance to Transgressors: Why Authorities Deliver Harsher Penalties When the Social Context Elicits Expectations of Leniency

Celia Moore[1]* *and Lamar Pierce*[2]

[1] Organisational Behaviour, London Business School, London, UK, [2] Olin Business School, Washington University in St. Louis, St. Louis, MO, USA

This paper combines experimental and field data to examine how authorities with discretion over how rules are enforced penalize transgressors when the social context of the transgression elicits expectations of leniency. Specifically, we test how transgressors are punished when it is their birthday: a day that triggers expectations of lenient treatment. First, in three scenario studies we explore individuals' intuitions about how they would behave and expect to be treated if they transgressed on their birthdays, as well as how they would imagine penalizing a birthday transgressor. Second, using more than 134,000 arrest records for drunk driving in Washington State, we establish that police officers penalize drivers more harshly when it is their birthday. Then, in a lab experiment in which we grant participants discretion over enforcing the rules of an essay-writing contest, we test psychological reactance toward transgressors who make their birthday salient, even subtly, as the mechanism behind this increased stringency. We rule out several alternative explanations for this effect, including public safety concerns, negative affect and overcompensation for bias. We conclude with a discussion of the theoretical and practical implications of our findings for the literatures on punishment, rule-breaking, and legal transgressions.

Keywords: ethics, transgressions, punishment, leniency, psychological reactance, drunk driving

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Marijke C. Leliveld,
University of Groningen, Netherlands
Hans Risselada,
University of Groningen, Netherlands
Janet Schwartz,
Tulane University, USA

***Correspondence:**
Celia Moore
cmoore@london.edu

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *10 September 2015*
Accepted: *04 April 2016*
Published: *09 May 2016*

Citation:
Moore C and Pierce L (2016) Reactance to Transgressors: Why Authorities Deliver Harsher Penalties When the Social Context Elicits Expectations of Leniency. Front. Psychol. 7:550. doi: 10.3389/fpsyg.2016.00550

INTRODUCTION

Scholars have theorized about when and how to punish individuals who transgress laws, rules, or regulations (Arvey and Jones, 1985; Butterfield et al., 1996), examined the consequences of punishment, in particular its impact on the attitudes and subsequent behavior of the punished individual (Ball et al., 1994; Podsakoff et al., 2006), and looked at how formal systems or written policies and procedures shape punishment decisions (Beyer and Trice, 1984). Other work has explored what motivates individuals to punish, and how their judgments of appropriate punishment change as a function of the seriousness of the offense and the intentions of the offender (Robinson and Darley, 1997; Carlsmith et al., 2002; Carlsmith and Darley, 2008). However, this body of work has overlooked how the social context of a transgression influences punishment decisions.

In this paper, we examine how transgressions are penalized when they occur in a social context that elicits expectations of leniency. Transgressions that occur in a social context in which the transgressor expects leniency put authorities with the discretion over punishing them into a difficult bind, needing to balance the motivation to meet the expectations elicited by the social context with the competing motivation to punish fairly and effectively. We argue that although transgressors may believe that transgressing in a social context that elicits expectations of leniency will lead to lighter penalties, this belief is misguided. Instead, we argue that—contrary to intuition—when authorities have a responsibility to enforce rules, but face a conflicting motivation to be lenient, they resolve this conflict in favor of harsher penalties rather than in favor of increased leniency.

Our research makes several theoretical contributions. First, we contribute to existing research on punishment by exploring how the social context of transgressions influences punishment decisions. Second, we extend our understanding of how individuals manage conflicting motivations when they have discretion over penalties. This is important because situational factors that trigger expectations of preferential treatment are pervasive, but many do not justify leniency in punishment decisions. Third, our work extends the literature on bias in punishment decisions by shifting the focus from discrimination on the basis of demographic characteristics such as race or gender to a focus on how people manage competing motivations to act. Ultimately, this work informs our understanding of the challenges in exercising discretion fairly and effectively (Kadish, 1961; Sherman, 1984; Pierce and Snyder, 2012).

DISCRETION IN PUNISHMENT

While laws and regulations provide guidelines for how to punish transgressions, individuals (e.g., managers, judges, or police officers), and groups (e.g., panels, boards, or juries) typically have discretion regarding whether and how much to punish those who transgress. Discretion over arrests (Reiss, 1984) and prosecutions (LaFave, 1970) is a central element of most legal and regulatory regimes because it allows authorities to consider an act's potential mitigating circumstances. However, discretion can also have negative consequences, including threats to due process (Kadish, 1961), abuses of power (Vorenberg, 1976), and biased treatment of individuals (Smith and Alpert, 2007). Whether discretion can be exercised appropriately is important, as exercising it poorly can delegitimize the work of authorities and undermine the equity and the efficacy of enforcement systems.

In the United States, the risk that discretion in punishment leads to unfair treatment of certain demographic groups has led to a number of high-profile initiatives to understand the extent and implications of these biases (e.g., Police Executive Research Forum, 2001; Lovrich et al., 2007). Most of these efforts have focused on ensuring that those with discretion over punishment do not treat transgressors differentially based on their demographic characteristics. Meanwhile, public discourse has neglected other potential biases affecting punishment decisions. Here, we suggest that individuals have difficulty managing situations where their formal responsibility to punish conflicts with a situational contingency that leads to expectations that a transgressor will be treated leniently.

Expectations of Leniency vs. an Obligation to Punish

One body of work that focuses directly on how we are motivated to punish is the literature on "just deserts" (Carlsmith et al., 2002; Carlsmith and Darley, 2008). This research explores how individuals shift their view of appropriate penalties for a crime depending on characteristics of the act and its perpetrators. The fundamental finding of this literature is that individuals are motivated to punish crimes in proportion to the magnitude of the harm they have caused and the availability of extenuating circumstances for the act. Though some aspects of the social context in which an offense occurs create extenuating circumstances, the only contextual factors that have been studied as motivations for leniency in punishment decisions are those that are directly relevant to attributions of blame or responsibility, such as whether the act was accidental (Carlsmith et al., 2002).

However, transgressions always occur in a broader social context. Elements of this broader social context may motivate expectations of leniency, but are arguably unrelated to the crime itself. Some elements of the social context create legitimate reasons to treat transgressors gently. For example, there is a norm of treating young people more leniently than adults when they transgress rules, as there are good arguments for why culpability is impaired before reaching maturity (Steinberg and Scott, 2003). Similarly, victims of long-term domestic violence and abuse are often punished with leniency, as their violent crimes are considered more justifiable (Ammons, 1994). These aspects of the social context that motivate expectations of leniency are often enshrined in legal structures, as evidenced in different sentencing guidelines for juvenile offenders, or special legal exceptions in the case of battered women.

Other aspects of the social context that motivate people to treat transgressors more leniently are less legitimate. Social norms of deference to authority (Milgram, 1974; Cialdini, 2009) result in higher-status individuals receiving more lenient penalties than lower-status individuals, at least for minor-to-moderate transgressions (Karelaia and Keck, 2012). The attractiveness of a perpetrator also appears to motivate more lenient punishment, even though how attractive someone is has nothing to do with how a transgression ought to be penalized (Sigall and Ostrove, 1975; Piehl, 1977; Stewart, 1985). These aspects of the overarching context of the crime clearly motivate more lenient treatment of these offenders, but there are strong arguments against using them as reasons for leniency.

"Special days" such as birthdays also elicit expectations of preferential or favorable treatment that may extend to expectations of leniency in the context of transgressions.

Birthdays are part of a larger class of days that have social or religious significance (e.g., Christmas, Yom Kippur) and are associated with strong norms of helping, kindness, and forgiveness. These days affect pro-social behavior such as charitable contributions (Jiobu and Knowles, 1974; Waldfogel, 1993). Birthdays specifically elicit expectations of favorable treatment, particularly for the individual whose birthday it is (Greene et al., 1987). The fact that many retailers offer free goods or services on individuals' birthdays, from pints of beer to pizzas, likely reinforces these expectations[1].

It is not a big leap to suggest that the expectation that individuals should receive special treatment on their birthdays will even extend to expectations of leniency when important rules are broken. On December 2, 2012, the rapper known as "The Game" was pulled over by Los Angeles Police because the car he was driving had invalid license plates. A celebrity website recounted that although the car was unregistered, the police released him and did not tow his car "because it was his birthday."[2] The reason the website reported for the leniency shown by the officers – "because it was his birthday" – suggests that the strength of the social cues to treat people favorable treatment on this day will extend to include leniency for legal transgressions.

In this paper, we focus on the social context of birthdays as a situational cue that motivates leniency for two reasons. First, even though this norm should be irrelevant in punishment decisions, for someone with the responsibility to penalize transgressors, a birthday elicits a competing motivation to treat that particular individual with leniency. Second, birthdays are ideal for studying the effect of the social context in non-experimental field settings because they are randomly distributed in the population: in most contexts, authorities that apprehend transgressors do not and could not have known it was transgressor's birthday before apprehending them. This means that birthday and non-birthday transgressors are randomly assigned to punishers, reducing endogeneity concerns about selection bias and causality when identifying variation in punishment decisions using data from the field.

Hypothesis 1a. Individuals will expect more lenient punishment in a social context associated with preferential treatment (i.e., individuals will predict that transgressors will be punished more leniently if it is their birthday).

The Transgressor's Perspective

If someone is caught transgressing a rule on their birthday, they have a choice between making that fact salient or not. There are several reasons why transgressors are likely to make salient a relevant fact that may motivate lenient treatment of them ("But it's my birthday!"). Individuals tend to volunteer reasons for their misbehavior in order to save face and reduce embarrassment (Goffman, 1955; Keltner and Anderson, 2000). Individuals also use available reasons in an effort to excuse their misbehavior and to transform how responsibility for actions are understood (Snyder et al., 1983). Given the strong expectation of favorable treatment associated with birthdays, we predict the following:

Hypothesis 1b. If an individual transgresses on their birthday, they will volunteer that fact in an effort to secure leniency.

The Authority's Perspective

The person in a position of authority needs to manage competing motivations to be lenient and to punish in a fair and effective way. Several reasons suggest that these competing motivations might be resolved in favor of increased leniency. Of course, the authority might try to ignore the expectation of leniency and punish transgressions as if this expectation did not exist. However, as prior research demonstrates, individuals often proceed rather automatically to enact scripted cues to behave in certain ways, even when the cue is completely irrelevant to the behavior it mindlessly triggers (Langer et al., 1978). Authority figures may be more lenient in this situation because they mindlessly enact this scripted cue.

Alternatively, it might be uncomfortable for an authority to behave counter to this expectation, particularly if they have the discretion over the transgressor's punishment. Extensive research shows that people tend to behave consistently with what is perceived to be the normative expectation in the situation, particularly when those norms are made salient (Reno et al., 1993). Thus, if an authority figure has discretion over the transgressor's punishment, their motivation to comply with social norms would also suggest they will treat the transgressor more leniently.

Such a prediction, however, ignores important psychological mechanisms involving the interaction between the transgressor and the authority. If the transgressor draws attention to his birthday, even subtly (as they are likely to, in an effort to capitalize on the expectation they have that this will lead to lenient treatment), this may shift the decision process of the authority. Although authorities may be indifferent or even positively inclined to treat a birthday transgressor with leniency, they may be particularly sensitive to perceptions that leniency is being solicited and react negatively to them. Consistent with the drive to punish in a fair and effective way, authorities may penalize transgressions more stringently when the social context cues expectations of leniency, because they will react negatively to any perception that their leniency is being solicited.

There is sound theoretical basis for such a prediction of stringency in psychological reactance theory (Brehm, 1966; Miron and Brehm, 2006). Three factors support the argument that authorities will experience psychological reactance when their obligation to punish occurs in a social context associated with expectations of leniency. First, individuals react strongly to sources of external influence they perceive as restricting their behavioral autonomy. In situations where individuals have the discretion to help others, any perception that their benevolence is not freely volunteered will trigger reactance. The level of reactance triggered is magnified to the extent that the freedom "not to help" is important (Brehm and Brehm, 1981, p. 171). Since those charged with penalizing transgressions are strongly

[1] http://www.mirror.co.uk/money/free-birthday-offers-deals-club-5260023

[2] http://www.tmz.com/2012/12/02/game-lapd-birthday-bentley-registration/

motivated not to help those whom they are obligated to punish, any perception that an individual is attempting to capitalize on aspects of the social context that trigger expectations of leniency will elicit reactance.

Second, reactance effects increase when requests for help appear inappropriate or illegitimate (Berkowitz, 1969, 1973). For example, Berkowitz (1969) found that individuals were less likely to help when they felt they were being coerced, and Gibbons and Wicklund (1982) suggested that acts of spontaneous helping require a situational cue to help that is both salient and legitimate. In other words, although a birthday might be a legitimate reason to let someone choose a restaurant that no one else likes, a birthday is an inappropriate reason to excuse him from the consequences of transgressing rules or laws. Thus, making a transgressor's birthday salient in the context of a transgression will likely elicit reactance.

Finally, when authorities have an obligation to penalize transgressions, they are motivated to ensure that the punishment is fair and appropriate. Brehm and Cole (1966) found that requests for help were counterproductive and elicited reactance when target participants were told to evaluate the person requesting help accurately. A motivation to treat someone fairly and appropriately is similar to a motivation to evaluate someone accurately. Thus, we argue that even a subtle perception that a transgressor is trying to use aspects of her social context to solicit more lenient treatment will trigger psychological reactance – because this will lead the authority to perceive that their freedom to exercise that discretion is being threatened.

Hypothesis 2. When a transgression occurs in a social context that elicits expectations of leniency, individuals with discretion over penalizing that transgression will react negatively to any action that makes this expectation salient (such as mentioning the fact that it is a transgressor's birthday).

We argue that when individuals perceive that someone is demanding something from them, whether the demand is explicit (actively soliciting leniency) or implicit (making salient an element of the social context that creates an expectation of leniency), they will experience the demand as a threat to their autonomy, and become less inclined to do it (Berkowitz, 1973). Regardless of the source of the perceived autonomy threat (e.g., choice restrictions, influence from norms, suggestions), individuals are motivated to counter the restriction and take actions to reestablish the threatened autonomy (Brehm, 1966). This reactance can operate below conscious awareness or intent (Chartrand et al., 2007) but can be extreme enough to cause a behavioral backlash in which the individual does the opposite of what she believes she is being asked to do (Fitzsimons and Lehmann, 2004). Thus, we propose:

Hypothesis 3a. Transgressions will be penalized more stringently when the social context in which the transgression occurs creates expectations of leniency (such as when it is the transgressor's birthday).

Hypothesis 3b. The increased stringency with which transgressions will be penalized when the social context elicits expectations of leniency will be mediated by psychological reactance.

OVERVIEW OF STUDIES

We now present six studies that test this argument using multiple methods and data sources. First, a series of scenario studies establishes that birthdays do represent a social context in which transgressors expect leniency, even though when individuals imagine themselves as authority figures with discretion over punishment decisions, they report higher levels of psychological reactance toward birthday offenders. Second, using 9 years of DUI (Driving Under the Influence) arrest records in the state of Washington (over 134,000 arrest records), we show that police officers punish marginal offenders more stringently on their birthdays than on other days. In a series of robustness checks, we show that it is unlikely that these results are explained by substantive differences in intoxication or public safety risk, but are instead likely based on the discretionary decisions of officers. Third, in a lab experiment in which we vary the birthday status of individuals who have transgressed rules, we demonstrate that individuals treat transgressors more stringently on their birthdays as a function of the psychological reactance triggered by the birthday status of the transgressor. A final study, using a similar experimental paradigm in the lab, rules out overcompensation for bias as the mechanism behind our effects.

Studies 1a–c: Individuals' Intuitions about Birthday Transgressions

We ran three studies, in separate online samples, to explore individuals' intuitions about whether they would expect to be treated leniently if they were pulled over for drunk on their birthday (Study 1a), whether they would volunteer that it was their birthday if they happened to be pulled over by a police officer on that day (Study 1b), and what they believe they would do themselves if they had discretion over penalizing a marginally drunk driver on their birthday (Study 1c). Together, our aim was to build a picture of what might occur in an actual interaction between a transgressor and an authority with discretion over penalizing the transgression on the transgressor's birthday. These studies were conducted on Amazon Turk, an online labor market where 'requesters' can post short tasks for 'workers' to complete for a small fee. Studies have found that data collected through Amazon Turk are of comparable quality to data collected through more traditional methods (Buhrmester et al., 2011; Goodman et al., 2013; Hauser and Schwarz, 2015).

Study 1a

The first solicited individuals' intuitions about how they expect they would be treated if they made their birthday salient in the context of transgressing. We predicted that their intuition would be that they would be treated more leniently on their birthdays.

Participants and procedure

We paid 306 participants (60% male; $M_{age} = 32$ years, $SD = 11.1$) \$0.50 to respond to a scenario. There were three conditions in the experiment. In a control condition, nothing about a birthday was mentioned. In two additional conditions, we asked participants to imagine it was their birthday, which they either mentioned to the officer [**birthday-mentioned**] or not [**birthday-not-mentioned**].

We included a birthday-not-mentioned condition to understand whether individuals' intuitions about their treatment would depend on whether or not they made their birthday salient to the officer. The scenario read:

> *Imagine you are driving home after an evening out with friends. You had a couple of drinks but you feel OK driving home by yourself. As you are driving, the local police stop you. The officer notices a faint smell of alcohol, though you are speaking clearly. To be safe, they ask you to take a breathalyzer test. It turns out that your blood alcohol content is 0.075%. The legal limit is 0.08%. Since your BAC is just below the legal limit, the local cops have discretion about how to proceed. While they are not required to arrest you, they may do so and test you again at the police station. They may also choose to release you with a warning.* [**birthday-mentioned:** *Imagine also that it is your birthday. You* [**birthday-not-mentioned:** ***do not***] *mention this to the police officer who has stopped you.*]

Participants were then asked to make a forced choice prediction about whether the officer would arrest them or release them with a warning.

Results

No one failed the attention check in this study, and everyone completed the main outcome measures; thus, results are reported for the whole sample. There were significant differences by condition in terms of the proportion of respondents who believed they would be arrested, $\chi^2(1, N = 306) = 9.45$, $p = 0.009$. When asked to predict what the officer would do, 25% of respondents in the birthday-mentioned condition and 35% of respondents in the birthday-not-mentioned condition believed they would be arrested, which represent more lenient treatment than the 45% of respondents in the control condition who predicted they would be arrested. Greater leniency was predicted in the two birthday conditions, compared to the control condition, $\chi^2(1, N = 306) = 7.21$, $p = 0.007$. The difference between the birthday-mentioned and birthday-not-mentioned condition was not statistically significant at conventional levels, $\chi^2(1, N = 209) = 2.43$, $p = 0.12$, although the results were consistent with greater expectations of leniency in the birthday-mentioned condition, compared to the birthday-not-mentioned condition. These results provide support for Hypothesis 1a, that individuals expect lenient treatment for transgressing when it is their birthday, particularly if they mentioned it.

Study 1b

This study solicited individuals' intuitions about they would do if they were stopped for drinking and driving on their birthday, as well as reasons behind their choice.

Participants and procedure

We paid 112 participants (56% male; $M_{age} = 34$ years, $SD = 10.5$) $0.50 to answer five questions and complete some basic demographic information. Participants read:

> *It is your birthday, and you've been out with friends celebrating. While driving home, you get pulled over by a police officer and asked to take a breathalyzer test. In your interactions with the driver, do you mention to the police officer that it is your birthday?*

We then asked them to indicate (on a 5-point scale) to what extent they agreed with four statements about why they might have made the choice they did: (1) It would result in the most lenient treatment from the officer; (2) It was the best excuse for my behavior; (3) It was the most appropriate choice to make; and (4) It would be the easiest thing to do.

Results

We did not include an attention check in this study, so results are reported for the whole sample. Thirty-five of the respondents (31%) said that they would mention their birthday to the officer. Those who said they would mention their birthday to the officer reported significantly higher levels of agreement with the statements that doing so: (1) would result in more lenient treatment from the officer [$M_{mentioned} = 3.31$, $SD = 1.02$ vs. $M_{not\ mentioned} = 2.05$, $SD = 0.83$, $t(110) = 6.94$, $p < 0.001$], (2) was the best excuse for their behavior [$M_{mentioned} = 3.20$, $SD = 1.16$ vs. $M_{not\ mentioned} = 1.83$, $SD = 0.79$, $t(110) = 7.33$, $p < 0.001$], and (3) would be the easiest thing to do [$M_{mentioned} = 3.69$, $SD = 0.90$ vs. $M_{not\ mentioned} = 2.78$, $SD = 1.19$, $t(110) = 4.02$, $p < 0.001$]. Both groups reported their choice was equally appropriate [$M_{mentioned} = 3.26$, $SD = 0.78$ vs. $M_{not\ mentioned} = 3.29$, $SD = 1.36$, $t(110) = 0.12$, $p = 0.91$]. These results provide some support for Hypothesis 1b. A substantial minority of individuals claim that they would mention it was their birthday to a police officer if they transgressed on their birthday. In addition, consistent with Hypothesis 1a, individuals who reported they would mention it was their birthday expected that doing so would lead to more lenient punishment for their offense.

Study 1c

In our final scenario study, we asked participants to imagine themselves in the role of the police officer. We wanted to see if the leniency they predicted they would receive as the driver would translate when they imagined themselves in the role of the police officer. We also wanted to assess how individuals in the role of the authority reacted to different ways that drivers might make their birthday salient, as a preliminary test of Hypothesis 2.

Participants and procedure

We paid 273 participants (62% male; $M_{age} = 32$ years, $SD = 9.8$) $0.50 to respond to a scenario. The experiment had four conditions: a control condition, and three birthday conditions (mentioned, soliciting-leniency, and noticed). We included several different birthday conditions to develop a more complete understanding of the outcomes of a range of possible interactions between the driver and officer. The scenario read:

> *Imagine you are a police officer conducting a road patrol. When you stop the next driver, you notice a faint smell of alcohol, though he is speaking clearly. To be safe, you require him to take a breathalyzer test. It turns out his blood alcohol content is 0.075%. The driver is under the 0.08% legal limit for Blood Alcohol Content (BAC), so you are not required to arrest him. However, you're concerned the breathalyzer test might not accurately reflect the impairment level of the driver, so you might want to arrest him as well.*

[**Control**] *As you consider your decision, he tells you that he is on his way home from dinner.*
[**Birthday-mentioned**] *As you consider your decision, he tells you that he is on his way home from dinner, and mentions that it is his birthday today.*
[**Birthday-soliciting-leniency**] *As you consider your decision, he tells you that he is on his way home from dinner, and mentions that since it is his birthday today, it would be nice for you to let him go with a warning.*
[**Birthday-noticed**] *As you consider your decision, he tells you that he is on his way home from dinner. As you take his driver's license back to your vehicle for some paperwork, you happen to notice that today is the driver's birthday.*

Participants were then asked to make a forced choice prediction about whether they would arrest the driver or release them with a warning.

We also tested individuals' psychological reactions to the scenarios. We used a 3-item measure of threat to freedom that has been used to study psychological reactance (Dillard and Shen, 2005). The items ("The driver tried to make my decision for me," "The driver was trying to manipulate me," and "The driver was trying to pressure me") were measured on a 5-point scale from strongly disagree to strongly agree ($\alpha = 0.91$). In addition, reactance theory suggested that threats to one's perceived autonomy might trigger "hostile and aggressive feelings" (Brehm, 1966, p. 9), though the theory claims that reactance may be present regardless of whether it is accompanied by such emotions. Dillard and Shen (2005) measured this type of negative affect using four items (irritated, angry, annoyed, aggravated), on a 5-point scale that ranged from "not at all" to "to a large extent" ($\alpha = 0.92$).

Results

Four participants failed the attention check question in this study; results are reported for the remaining 269 participants. There were significant differences in whether participants reported they would arrest the driver, $\chi^2(3, N = 269) = 12.95$, $p = 0.005$. In the control condition, 21% said they would arrest the driver, which was not significantly different from the 16% who said they would arrest the driver in the birthday-mentioned condition, $\chi^2(1, N = 135) = 39$, $p = 0.53$, nor the 12% who said they would arrest the driver in the birthday-noticed condition, $\chi^2(1, N = 135) = 1.85$, $p = 0.17$. However, when the driver mentioned it was his birthday in an effort to solicit leniency, individuals were significantly more likely (36%) to predict they would arrest the driver, $\chi^2(1, N = 135) = 3.87$, $p = 0.049$ than in the control condition.

The scenarios also elicited different levels of psychological reactance in the participants, $F(3,265) = 35.41$, $p < 0.001$. Results showed a significant linear trend, $F(1,265) = 83.56$, $p < 0.001$, such that participants were significantly more likely to perceive a threat to their freedom as the driver made the birthday increasingly salient. The least reactance was reported in the control condition ($M = 1.95$, $SD = 0.93$) and the birthday noticed condition ($M = 1.74$, $SD = 0.93$), which did not differ from each other ($p = 0.22$). This level rose significantly in the birthday-mentioned condition ($M = 2.68$, $SD = 0.93$, $p < 0.001$) and again in the birthday-soliciting-leniency condition ($M = 3.30$, $SD = 1.07$, $p < 0.001$). The difference between the birthday-mentioned and birthday-soliciting-leniency conditions was also significant ($p < 0.001$).

Participants' negative affect also significantly differed by condition, $F(3,265) = 5.44$, $p < 0.001$. However, the birthday-soliciting-leniency condition ($M = 2.24$, $SD = 1.06$) was the only condition that significantly differed from the rest (all at $p < 0.001$), which were statistically indistinguishable from each other ($M_{\text{birthday-mentioned}} = 1.70$, $SD = 0.86$; $M_{\text{birthday-noticed}} = 1.63$, $SD = 0.83$; $M_{\text{control}} = 1.69$, $SD = 0.81$).

These results provide preliminary support for Hypothesis 2, that a transgressor's birthday elicits negative psychological reactions among individuals with discretion over their punishment. In addition, the more obvious the effort to capitalize on the social expectation of leniency, the more negative the reaction.

Discussion

Together, these results suggest three things. First, birthdays do represent a social context in which individuals expect to receive lenient treatment for their transgressions – even if the transgression is quite severe. Second, though still a minority, a substantial proportion of individuals claim they would mention it was their birthday to a police officer if they were pulled over for drunk driving on that day. Third, individuals imagining themselves in the role of a police officer believe they would only treat birthday offenders more stringently if the driver attempted to use that fact to solicit lenient treatment for his offense. Third, even though respondents reported they would only treat birthday offenders more stringently if they used the birthday to solicit leniency (this was also the only condition that elicited significantly more negative affect from the respondent), any mention of the transgressor's birthday elicited psychological reactance. This last finding suggests that individuals with discretion over punishment may have more general psychological reactions to birthday transgressors.

Study 2: Field Evidence from Drunk Driving Stops by Officers

In Study 2, we use a unique sample of field data to identify how individuals in positions of authority actually penalize transgressions in a social context that elicits expectations of leniency. Specifically, we study arrests involving suspicion of DUI of alcohol in the state of Washington, and test whether otherwise similar drivers are more likely to be arrested if it is their birthday, compared to those for whom it is not their birthday.

Empirical Context

In all U.S. states, driving while intoxicated by alcohol (drunk driving) is prohibited and has a severe impact on public safety. Economists have estimated that intoxicated drivers create externalities of at least 30 cents per mile driven due to social welfare costs of traffic fatalities (Levitt and Porter, 2001). Alcohol-related fatalities in the United States were estimated to be 11,948 in 2010, representing 36% of all traffic fatalities

that year (National Highway Traffic Safety Administration, 2010). Furthermore, deterring drunk driving is difficult, with estimates that only one out of every 2,000 drunk drivers is actually arrested (personal communication, Washington State Patrol).

Driving under the influence laws are enforced by several police agencies in Washington State, including the Washington State Patrol, which is responsible for monitoring and enforcing the state's highway systems, as well as local agencies, including municipal police, county sheriff's offices, and Indian Nations agencies. In Washington State, DUI laws are primarily based on the driver's blood alcohol level (BAC). Drivers whose BAC exceeds 0.08% are said to be in *per se* violation of state law and have little legal defense. Such drivers face minimum penalties of $865, 24 h incarceration, and 90 days suspended license for their first offense. Drivers with BAC levels above 0.15% are subject to even greater penalties, including minimum fines of $1,120, 2 days incarceration, and a 1-year revocation of one's driver's license. Penalties escalate rapidly with repeat offenses. **Figure 1** presents the average relationship between drinking behavior and BAC, conditional on gender and body weight, though food consumption, regular alcohol consumption, and genetic factors also influence BAC. As the body processes alcohol, BAC drops at an average rate of 0.015 per hour.

When an officer suspects a driver of DUI, she typically administers a field sobriety test. Furthermore, the officer administers a mobile breath test ("breathalyzer"), which estimates the BAC of the driver. If the officer determines the driver to be intoxicated, the driver is placed under arrest and taken to a field station for a formal (and admissible in court) breath test. If the officer observes a mobile BAC greater than 0.08, the decision is straightforward. The driver is almost certainly in *per se* violation, and the officer arrests the driver. However, the decision is much less clear if the mobile BAC is below 0.08. When the mobile BAC is below 0.08, the officer has discretion over whether or not to arrest the driver. Drivers with BAC levels between 0.04 and 0.079, for example, are likely impaired, but less so than *per se* violators. These "marginal offenders" are arrested at the discretion of the officer. We use the term "marginal" to refer to drivers who fall just below the *per se* blood alcohol threshold[3].

Arresting the driver presents several potential costs for the officer. First, the arrest process takes the officer off the road for several hours and thereby precludes her from potentially arresting an even more highly intoxicated driver. Second, drivers who do not violate the *per se* rule are much more difficult to prosecute, as a conviction must rely on the officer's evaluation of the driver's intoxication. Consequently, prosecuting attorneys typically discourage officers from arresting drivers with low BAC, and most of these cases are plea-bargained (decided without going to court) with minimal penalties.

Data

Our data include every DUI arrest in Washington State from 2001 to 2009. These data include the agency and identity of the arresting officer as well as the name, age, gender, and ethnicity of the driver. Also included are the date, time, and location of the arrest. The data also note the primary criminal charge, which allows us to exclude DUI arrests that are secondary to more severe crimes such as weapons violations, violent crimes, or outstanding arrest warrants. The data also identify the mobile BAC reading, when taken, as well as the court-admissible BAC reading from the police station. Since the data also identify the exact time of each test, we know the length of delay before the driver was given the court-admissible test. We present basic summary statistics in **Table 1** for both the pooled sample as well as the sample separated by birthday/non-birthday. The average BAC for all arrests is 0.13, with 94% above the *per se* threshold. Approximately one out of every 300 arrests is a birthday driver. The average age of

[3]We use this terminology to indicate that such drivers are at the *per se* margin, not to understate the danger or seriousness of driving at these BAC levels. Driving with a marginal BAC of 0.07, for example, still elevates the risk of injury and fatality considerably, and we intend no judgment on the ethicality (or lack thereof) of such behavior.

A

drinks	body weight (lbs)								
	100	120	140	160	180	200	220	240	
0	0	0	0	0	0	0	0	0	Only Safe Limit
1	.05	.04	.03	.03	.03	.02	.02	.02	.01-.03 Impairment begins
2	.09	.08	.06	.06	.05	.05	.04	.04	
3	.14	.11	.10	.09	.08	.07	.06	.06	.04-.07 Driving skills deteriorating; you can be arrested for DUI
4	.18	.15	.13	.11	.10	.09	.08	.08	
5	.23	.19	.16	.14	.13	.11	.10	.09	
6	.27	.23	.19	.17	.15	.14	.12	.11	.08 Illegal to drive, immediately lose license; subject to criminal penalties, fines and/or jail.
7	.32	.27	.23	.20	.18	.16	.14	.13	
8	.36	.30	.26	.23	.20	.18	.17	.15	

B

drinks	body weight (lbs)									
	100	120	140	160	180	200	220	240	260	
0	0	0	0	0	0	0	0	0	0	Only Safe Limit
1	.04	.03	.03	.02	.02	.02	.02	.02	.01	.01-.03 Impairment begins
2	.07	.06	.05	.05	.04	.04	.03	.03	.03	
3	.11	.09	.08	.07	.06	.06	.05	.05	.04	.04-.07 Driving skills deteriorating; you can be arrested for DUI
4	.15	.12	.11	.09	.08	.07	.07	.06	.06	
5	.19	.16	.13	.12	.10	.09	.08	.08	.07	
6	.22	.19	.16	.14	.12	.11	.10	.09	.09	.08 Illegal to drive, immediately lose license; subject to criminal penalties, fines and/or jail.
7	.26	.22	.19	.16	.15	.13	.12	.11	.10	
8	.30	.25	.21	.19	.17	.15	.14	.12	.12	

FIGURE 1 | Relationship between Alcohol Consumption and BAC. (A) Represents the average blood alcohol content for women based on number of drinks (12 oz. beer, 5 oz. wine, 1.5 oz. hard alcohol) and body weight. **(B)** Represents men.

TABLE 1 | Study 2: Descriptive statistics for DUI arrests.

	All arrests		Birthday arrests	Other arrests
Variable	**Mean**	***SD***	**Mean**	**Mean**
Field BAC	0.133	0.049	0.134	0.134
Per se violation	0.94	0.24	0.92	0.94
Field BAC – Station BAC	−2.18	35.12	−1.06	−2.18
Minutes from field to station	60.73	46.75	59.17	60.74
Birthday driver	0.004	0.062	1	0
Driver age	34.33	11.38	37.08	34.32
Female driver	0.21	0.41	0.23	0.21
White driver	0.83	0.38	0.82	0.83
Number of observations	134,507		518	133,989

the sample is 34, 21% are female and 81% are ethnically white (non-Hispanic).

One weakness in our data is that we are unable to observe drivers stopped for suspicion of DUI but not arrested. Only drivers who were arrested appear in our data, creating potential survivor bias in any standard regression analysis. We will address this weakness by exploiting the discrete threshold at BAC = 0.08 in order to infer distributions of non-arrested drivers in the data. Another weakness is the relative rarity of birthdays, which represent 0.38% of all arrests vs. 0.27% (one out of 365.25) of all days. This rarity means that we must infer differences in birthday traffic stops from a substantially smaller sample than the total DUI database.

Identification Strategy

Using driver's birthday as the context in which there is a social expectation of lenient treatment has several important characteristics from an identification perspective. First, the norm of preferential treatment on one's birthday is universally known and widely observed. Second, birthdays are unobservable to an officer prior to a traffic stop and thus unlikely to create an unobservable selection bias in traffic stops. Third, birthdays are randomly distributed and uncorrelated with other factors that might affect officer leniency. This third point is critical for our decision to examine birthdays instead of other holidays such as Valentine's Day, Mother's Day, or Christmas, which may affect the officer. An officer showing leniency on Valentine's Day, for example, may simply want to avoid a 2 h arrest that keeps him from dinner with a spouse, or he may be in a foul mood due to working on a holiday.

We identify officer stringency in DUI enforcement by observing how often officers arrest *per se* offenders relative to marginal offenders. While all officers must arrest *per se* offenders, extremely stringent enforcement would entail an increase in arrested marginal offenders relative to *per se* arrests. Officers may be able to identify extremely intoxicated drivers (e.g., BAC > 0.15) before a traffic stop, but it is unlikely they would be able to *ex ante* distinguish between marginal offenders and those with BAC levels just above the *per se* limit. Consequently, the ratio of traffic stops that involve BACs just above the threshold (e.g., BAC = 0.08) should be approximately equal to the frequency involving BACs just below (e.g., BAC = 0.079), as should the appearance of intoxication when the driver is first confronted. Given the approximately equal number of marginal and *per se* violators stopped and tested, the relative frequency of arrest of marginal offenders relative to borderline *per se* offenders is unlikely to reflect the choice to stop drivers and instead will reflect the decision to punish marginal offenders. This approach is similar to one recently used to examine possible racial bias in DUI stops (Horn et al., 2014).

Results

We first present birthday arrest frequency for marginal and *per se* violators for four different bandwidths surrounding the 0.08 *per se* threshold (see **Figure 2**). The white bar represents marginal offenders, while the gray bar reflects *per se* violators. Whiskers reflect plus or minus one standard error. The four decreasing bandwidths are represented from left to right, with the furthest left group indicating plus or minus 0.04 and the furthest right group representing plus or minus 0.01. **Figure 2** shows a much higher level of birthday arrests for marginal offenders than *per se* offenders, which suggests that discretion leads to increased stringency for birthday drivers stopped by police. Together, these results provide support for Hypothesis 3a, that birthday drivers receive increased stringency rather than increased leniency.

Regression analysis

The goal of our analysis is to identify how the behavioral interaction between the transgressor (driver) and punisher (officer) are affected by the expectation of leniency associated with birthdays. Consequently, we are concerned that the increased likelihood of a discretionary arrest on birthdays might simply reflect fundamental differences between the characteristics of birthday and non-birthday drivers. Similarly,

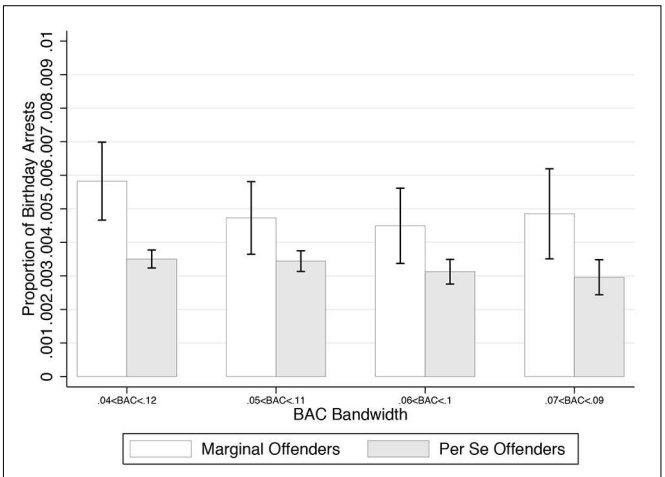

FIGURE 2 | Study 2: Proportion of arrests of birthday drivers for different bandwidths near the *per se* threshold. Marginal offenders represent those drivers with BAC below the 0.08 threshold. *Per se* offenders are above the threshold. Bars represent ±1 SE.

the police who arrest them or the conditions under which they are arrested might be different. To address this, we implement regression analysis that estimates a decrease in the likelihood of a birthday driver for all arrests above the *per se* threshold. This approach is similar to a regression discontinuity design, which involves estimating the impact of a discrete threshold in a continuous independent variable on an outcome variable (Imbens and Lemieux, 2008; Snyder, 2010; Pierce and Snyder, 2012; Pierce et al., 2013). We cannot achieve the standards of a true regression discontinuity design, because the extremely rare occurrence of birthday arrests does not provide sufficient observations very close (i.e., BAC values 0.079 and 0.08) to the *per se* threshold. We therefore urge caution in interpreting any causal relationship from our data.

Because officers cannot observe on which side of the threshold a moderately intoxicated driver lies before arrest, the assignment near the threshold is random for those stopped for DUI. Since DUI stops are randomly assigned to either side of the threshold, our theoretical argument is that the behavioral interaction between the driver and officer is the mechanism driving any discrete increase in birthday probability at the *per se* threshold. Since this mechanism is difficult to directly identify in the arrest data, the purpose of our regression model is to provide evidence that this difference is not due to observable driver, officer, or arrest characteristics that are correlated with birthdays but different from our argued mechanisms.

Our first specification uses logistic regression to estimate the probability that an arrest involves a driver birthday as a function of the *per se* rule. If officers treated birthday drivers identically to other drivers, we should expect arrests immediately on each side of the threshold to have equal probability of involving a birthday. Alternatively, if officers are more aggressive in punishing marginally drunk drivers on their birthday, we should expect a higher probability of a birthday for BAC < 0.08 and therefore a negative coefficient for the *per se* threshold. It is important to account for the underlying relationship between the dependent variable (birthday) and the continuous variable that defines the discrete threshold (BAC). We include a quartic polynomial of BAC as a control variable to allow for functional flexibility in the relationship between drinking behavior and birthdays.

Our base model with no control variables is presented in column 1 of **Table 2**, with logit coefficients and robust standard errors clustered at the officer level in parentheses. Column 2 adds flexible time controls, and column 3 adds controls for driver age (quartic polynomial), gender, and ethnicity. Also included are dummy variables for each county. Each column shows a negative relationship between the *per se* threshold and the probability of an arrestee birthday. The correct interpretation for these results is that the probability of an arrestee birthday distinctly drops when the BAC level crosses the threshold at 0.08% (thus removing the officer's discretion). Providing support for Hypothesis 3a, these models suggest that marginally drunk drivers who are stopped are more likely to be arrested on their birthday than on other days. To aid interpretation, we calculate the marginal effects for the fully controlled model, which are 0.0018 ($p = 0.06$). Given the base rate of birthday arrestees of approximately 0.4% of *per se* violators, one's probability of arrest

TABLE 2 | Study 2: Regression models predicting birthday arrests.

	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	OLS
Driver sample:	**All**	**All**	**All**	**All**
Dependent variable:	**Birthday**	**Birthday**	**Birthday**	**Birthday**
Per se violator	−0.543* (0.265)	−0.536* (0.265)	−0.488† (0.259)	−0.0021 (0.0013)
BAC	0.008 (0.013)	0.009 (0.013)	0.006 (0.013)	0.00003 (0.00005)
BAC^2	−0.00003 (0.0001)	−0.00004 (0.0001)	−6.3e-06 (1.3e−04)	−3.1e−08 (5.5e−07)
BAC^3	5.3e−08 (5.4e−07)	7.5e−08 (5.5e−07)	−3.8e−08 (5.4e−07)	−4.7e−10 (2.1e−09)
BAC^4	−8.4e−11 (7.4e−10)	−1.1e−10 (7.5e−10)	2.8e−11 (7.3e−10)	8.4e−3 (2.7e−12)
Month/Day dummies	No	Yes	Yes	Yes
Year dummies	No	Yes	Yes	Yes
Age	No	No	−0.363† (0.217)	−0.028** (0.003)
Age^2	No	No	0.012† (0.006)	0.001** (0.0001)
Age^3	No	No	−0.0002† (0.0001)	−0.00002** (1.7e−06)
Age^4	No	No	9.1e−07* (4.1e−07)	9.4e−08** (9.3e−09)
Male	No	No	−0.060 (0.112)	−0.00007 (0.0005)
Driver ethnicity dummies	No	No	Yes	Yes
County dummies	No	No	Yes	Yes
Officer FE	No	No	No	Yes
Pseudo R-squared	0.0007	0.0370	0.0412	0.0451
Number of observations	134,507	134,507	133,795	134,507

*Standard errors clustered at the officer level in parentheses. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$. Observations across models are not equal due to maximum likelihood estimation dropping perfectly predicted groups. Fixed effects in Model 4 are at the officer level. p-value for Model 4 is 0.11. Small coefficients and standard errors listed in scientific notation.*

increases by about 50% near the 0.08 BAC threshold on one's birthday.

As a robustness test, in column 4 we report a linear probability model with officer fixed effects, since the rarity of birthday events does not allow the use of logit models with fixed effects. The linear fixed effect model produces a coefficient very similar to our marginal effects, but with reduced statistical significance ($p = 0.11$), which is unsurprising given the coefficient is only identified off the smaller set of officers with at least one birthday arrest. Still, this model suggests that our results cannot be explained by the most stringent officers stopping birthday drivers. Other multi-level models that might explore agency- or officer-level predictors of stringency cannot be estimated, because few officers or agencies experience more than one or two birthday arrests.

We note that we cannot estimate the marginal effect at other points farther below the threshold due to our identification strategy. Furthermore, the low number of birthday arrests in our data makes more formal testing of regression discontinuity models difficult, so our results would be more compelling if we could triangulate them using additional data.

To address possible differences in our non-birthday and birthday samples, we also created a matched sample based on observable driver demographics, BAC, and stop characteristics. We implement a propensity score matching algorithm that chooses the ten nearest non-birthday neighbors for each birthday arrest, which reduces our sample to 5,117 arrests (some non-birthday arrests are neighbors for multiple birthday arrests). Using this matched sample, we repeat the *t*-tests and regressions reported in **Figure 2** and **Table 2**; these produce nearly equivalent results. The proportion of birthday offenders remains higher for marginal offenders for each of the four bandwidths used in **Figure 2** ($p < 0.01$), and the birthday coefficients for the three models presented in **Table 2** are very similar, despite a 96% decrease in sample size: −0.523 ($p = 0.06$), −0.496 ($p = 0.08$), −0.482 ($p = 0.09$).

Robustness tests

Our evidence of higher stringency toward birthday drivers supports Hypothesis 3a, but raises a number of alternative explanations. One natural concern with our identification strategy is that BAC readings may not accurately represent the public safety risks of birthday drivers, and that the increased stringency we observe represents a rational police response to expectations of future accidents or increasing intoxication. We systematically examine these alternative explanations by testing differences in arrested drivers' characteristics.

We first address whether the mobile BAC of birthday drivers stopped by police accurately reflects their level of alcohol consumption, relative to other drivers. In this alternative explanation, marginally drunk birthday drivers have alcohol in their stomach that has not yet entered the bloodstream due to binge drinking or a stomach full of food, and officers arrest the driver because of their tacit knowledge that their BAC will continue to climb above the *per se* threshold later in the evening. In such a case, the decision to disproportionately arrest marginally drunk birthday drivers would be rational and show great foresight. To examine whether birthday drivers are more likely to increase in BAC due to pre-stop drinking patterns, we examine the change in BAC between the mobile and station tests. This change reflects how much the driver's BAC increased or decreased between arrest and arrival at the testing facility during a time where additional drinking was not possible. The differences between average BAC changes of birthday (−0.0011) and non-birthday (−0.0022) drivers are indistinguishable ($p = 0.50$), as are the number of minutes between the two tests (60.4 vs. 59.1, $p = 0.51$). This suggests that BAC measures reflect equal intoxication of birthday and other drivers.

We next address the alternative explanation that marginally drunk birthday drivers may be inherently more dangerous than their non-birthday counterparts and that stringency toward them is a rational public safety response. For this alternative explanation, we test whether birthday drivers are more likely to drink (and drive) later in the evening, making their arrest a pre-emptive strategy for law enforcement. The average time of arrest is also nearly identical between the two groups (10:54 p.m. vs. 10:59 p.m., $p = 0.58$), suggesting that officers are not preemptively arresting birthday drivers earlier in the evening to avoid later drinking and driving.

We also examined whether arrested birthday drivers were more likely to have it be their first DUI arrest (in the state) compared to other drivers. To do so, we used only those drivers where no officer discretion was involved (thus eliminating any birthday bias), and found that although it was somewhat more likely that birthday drivers were being arrested for the first time compared to non-birthday drivers (91% vs. 89%), the difference was statistically indistinguishable (Fisher's exact test, $p = 0.19$).

Another alternative explanation is that officers might punish marginally drunk birthday drivers more stringently because they believe birthday drivers are more likely to be "scared straight" by the arrest. Although we cannot observe other confounds (such as differences in conviction and sentencing), we tested whether those whose first arrest was on their birthday were less likely to be arrested for a later DUI. For *per se* violators (where discretion was not involved), birthday drivers were slightly less likely to reoffend (17% vs. 19%, Fisher's exact test, $p = 0.11$), but there is virtually no difference among the lowest *per se* offenders who best approximate marginal offenders (BAC between 0.08 and 0.12). Birthday and other drivers both reoffended at a 17% rate ($p = 0.99$). These suggest that there is no strong deterrence reason why officers should arrest birthday offenders more often than other offenders. Even if they are, it is not effective. Marginal birthday offenders are, if anything, more likely than others to reoffend after being arrested (30% vs. 17%, Fisher's exact test, $p = 0.10$).

Finally, we examine whether intoxicated birthday drivers were more likely to be involved in an accident compared to other drunk drivers. Arresting after an accident where the driver has a positive BAC involve no police discretion (hence we had excluded arrests involving accidents from our main sample). To test whether officers may be arresting birthday drivers at a higher rate because they have insider knowledge that they

are more likely than other drivers to cause later accidents, we compare the ratio of birthday drivers among those arrested for DUI offenses ending in accidents (arrests excluded from our main sample) to the ratio of birthday drivers among discretionary DUI arrests. The percentage of drunk-driving accidents involving birthday drivers is 0.43%, compared to 0.41% for discretionary officer arrests ($p = 0.39$), suggesting that birthday drivers are no more likely to get into DUI accidents than drivers on other days.

This similarity in birthday rates for non-discretionary accident rates also casts doubt on an alternative explanation that officers give fewer breath tests to birthday drivers, and consequently might show more stringency toward those under 0.08 to compensate for this prior leniency. If that were the case, then we would expect a lower average rate of birthday drivers in discretionary tests (non-accidents) than in mandatory ones, which we do not. Together, these tests cast doubt on alternative explanations for police stringency toward birthday drivers, but of course cannot disprove them.

Discussion

Of course, we cannot know whether drivers are actually soliciting leniency in their interactions with police officers when they are pulled over on their birthdays. However, if Study 1b is any indication, a substantial minority of the individuals (31% of the study sample) reported that they would mention their birthday to the police officer. Alternatively, contrary to Study 1c, which suggested that reactance in the condition in which participants noticed the birthday was equal to the control condition, officers in the field may react negatively to drunk driver even if their birthday isn't mentioned. Whatever occurs between the officers and the drivers in the field, our analysis provides support for Hypothesis 3a: drivers who are at the margins of the legal limit for blood alcohol are more likely to be arrested on their birthday than on other days. This effect appears to be unrelated to the public safety risk of these drivers, their demographics, and the conditions under which they are arrested.

Study 3: Testing Psychological Reactance as a Mechanism in the Lab

The data from Study 2 do not allow us to test whether psychological reactance explains the apparent stringency toward birthday drivers, neither can they reveal whether this is a more general behavioral response or whether it is idiosyncratic to the setting of drunk driving. We address these concerns by designing a laboratory experiment using a different type of transgression, which additionally allows us to test psychological reactance as our hypothesized mechanism (Hypothesis 3b).

Participants and Procedure

The behavioral lab at a UK-based business school (43% male; $M_{age} = 28.5$ years, $SD = 9.7$) recruited 162 participants to complete the study for a £10 payment. The study was approved by the school's Ethics Review Board, and met all APA requirements for the ethical treatment of research participants.

Participants were randomly assigned to one of three conditions. There were two birthday conditions: one in which the transgressor was using his birthday as a reason to solicit preferential treatment (**birthday-soliciting-leniency**), and one in which the transgressor merely mentioned it was his or her birthday (**birthday-mentioned**). A control condition made no birthday reference.

We informed participants that the lab was partnering with a nearby school specializing in English as a Second Language to evaluate a student essay-writing competition. Participants were all assigned to the role of "evaluator" and tasked to judge three of the essays competing for prizes. We also told participants that, because teachers typically know the students in their classes before grading any of their work, the students had written a short paragraph about themselves, which would be attached to each essay. We used actual example essays from the American College Testing writing assessment arguing in favor of extending high school by 1 year. We chose two essays that the assessment service used as examples of poorly written essays (that had scored 1 and 2 out of 5) and one example of a good essay (that had scored 5 out of 5).

We provided participants with a scoring sheet and contest rules, which included a rule forbidding essays over 500 words from being eligible for prizes. The rule read**: "*The students were instructed to follow a 500-word limit. You should still grade their essay if it is more than 500 words, but if they exceed 500 words, it is ineligible for the prize.*"** They were instructed to judge each essay and asked whether they nominated any of the essays for either the first prize (a 10% tuition fee refund), or an honorable mention (a new backpack with the school logo). Finally, they were told that the students were aware that the essays were being evaluated by outside graders and that competition winners, chosen by them, would be announced "this coming Friday." Instructions stressed that the competition had meaningful outcomes for the students and that it was important for them to take their job seriously. Each participant evaluated the same three essays; however, the handwritten personal statements stapled to the essays varied. There were three versions: one written by a Brazilian female, one by a Mexican male, and one by a Spaniard whose gender was not made explicit. Personal statements were counterbalanced to ensure that any differences in participants' evaluations or prize nominations were unrelated to the personal messages' content, other than the birthday manipulation.

The birthday manipulation was included at the end of the personal statement attached to Essay #3 (the essay assessed by the American College Testing service as the one of the highest quality), which was always positioned last in the package. The handwritten personal statement also included a message that either mentioned the essay-writer's birthday [**birthday-mentioned**: "*It's my birthday next Friday, and I will be 22 years!*"], or suggested that the essay-writer deserved the prize because it was their birthday [**birthday-soliciting-leniency**: *"I really think I deserve the prize because it will be my birthday the day the prizes are announced—I will be 22 years!*"]. In the control condition, nothing was mentioned about the essay-writer's birthday. This manipulation allows us to test whether the participants who appeared to solicit leniency because it was their birthday would be penalized more harshly, or

whether simply mentioning the birthday would be enough to elicit the stringency effect we observed in the drunk driving data.

Measures

Participants were instructed to grade the essay's unique ideas (10 points), persuasiveness (10 points), language quality (10 points), and grammar, spelling and punctuation (10 points). These points were summed to create a total score. We used participants' scores as a manipulation check to confirm that the essay containing the birthday manipulation was evaluated as the "best" essay among the three, and thus the most likely to be nominated for a prize *if* the essay writer was not penalized for breaking the word limit rule.

Mechanism: psychological reactance

We measured participants' psychological reactance to each of the essay writers' personal statements using the same 3-item measure of threat to freedom used in Study 1c (Dillard and Shen, 2005). We measured the items for each essay writer ($\alpha = 0.83$ for essay writer 1, $\alpha = 0.86$ for essay writer 2, and $\alpha = 0.82$ for essay writer 3). To rule out the alternative explanation that negative affect (anger or annoyance) was driving our effects, we also included the same measure of negative affect used in Study 1c ($\alpha = 0.82$ for essay writer 1, $\alpha = 0.86$ for essay writer 2, and $\alpha = 0.82$ for essay writer 3).

Dependent variable: stringency

The word counts of each essay were handwritten on each of the essays and circled. At 513 words, Essay #3 violated the 500-word limit rule by 13 words. Neither of the other two essays violated the word limit. The dependent variable of interest was whether participants treated Essay #3 with increased stringency by not nominating it for the prize even though it was the best of the three essays.

Results

Five participants failed to complete all relevant measures, and were excluded from the analysis. Results are reported for the remaining 157 participants.

A repeated measures ANOVA with final score as the within-subjects factor confirmed the ranking of the essays provided by the American College Testing service. The essay scores significantly differed from each other, $F(2,155) = 326.55$, $p < 0.001$, and the score for Essay #3 ($M = 32.0$, $SD = 5.9$) was significantly higher than the scores for Essay #1 ($M = 23.1$, $SD = 5.6$) and Essay #2 ($M = 15.4$, $SD = 6.6$) scores. Thus, we interpret the failure to nominate Essay #3 for the prize as evidence that evaluators were penalizing this student for transgressing the word limit rule, effectively disqualifying the writer from the competition, rather than as evidence that the evaluator believed the essay to be low quality.

We next established that participants' psychological reactance to the third essay was affected by the condition to which they were assigned, $F(2,154) = 3.98$, $p = 0.021$. Consistent with Study 1c, this pattern followed a significant linear trend, $F(1,154) = 7.95$, $p = 0.005$. The highest levels of psychological reactance were felt by participants in the birthday-soliciting-leniency condition ($M = 2.83$, $SD = 1.14$), with slightly lower levels by participants in the birthday-mentioned condition ($M = 2.58$, $SD = 0.91$), and lowest levels in the control condition ($M = 2.27$, $SD = 0.96$). Participants clearly reacted more strongly as messages reflected more explicit attempts to capitalize on expectations that they would be treated preferentially on their birthday. Participants' levels of negative affective reaction to the third essay did not differ by condition, $F(2,154) = 0.11$, $p = 0.89$, $M_{birthday-mentioned} = 1.21$, $SD = 0.47$; $M_{birthday-soliciting-leniency} = 1.26$, $SD = 0.53$; $M_{control} = 1.25$, $SD = 0.58$. However, we note that our threat to freedom measure correlates with our measure of negative affect ($r = 0.44$, $p < 0.001$), indicating that a threat to freedom is experienced, in part, as negative affect.

Hypothesis 3b predicted that an authority figure's increased stringency (in this case, penalizing those who violated competition rules) as a function of targets' birthdays is driven by psychological reactance. We used Preacher and Hayes' PROCESS macro (Hayes, 2013) to test psychological reactance as the mediator in the relationship between transgressors' birthday statuses and whether they were denied the prize. Our design uses a dichotomous outcome variable and a multi-categorical independent variable. To test our predicted relationships, we constructed dummy variables for each condition (birthday-soliciting-leniency, birthday-mentioned, and control). For each model, one dummy variable is specified as the independent variable and a second dummy variable is included as a covariate; the resulting test of the indirect effect represents the comparison between the condition specified as the independent variable and the reference condition (excluded from the analysis). The macro generates bias-corrected bootstrap confidence intervals for each indirect effect.

We ran three models for all the relevant comparisons, each using 5,000 bootstrap samples. We included the participants' reactance toward the first and second essays as covariates in the analyses, as the reactance measures were significantly correlated with each other (between Essay 1 and Essay 2, $r = 0.68$, $p < 0.001$, between Essay 1 and Essay 3, $r = 0.37$, $p < 0.001$; and between Essay 2 and Essay 3, $r = 0.39$, $p < 0.001$), and we wanted to ensure that our models used the reactance triggered by our birthday manipulation as the mediator of our effects, rather than the reactance the essay writers' messages elicited overall. Results for all three models are reported in **Table 3**. Compared to the control condition, the indirect effect of either birthday condition on increased stringency via psychological reactance was positive, with 95% confidence intervals that excluded zero, indicating significant indirect effects via reactance. The indirect effect was largest comparing the birthday-soliciting-leniency condition to the control condition (point estimate = 0.31, 95% CI 0.035 to 0.709). The 95% confidence internal for the birthday-mentioned condition compared to the control condition also excluded zero (point estimate = 0.15, 95% CI 0.011 to 0.424). The birthday condition in which the student explicitly solicited leniency also showed a bigger indirect effect compared to the birthday-mentioned condition (point estimate = 0.16, 95% CI

TABLE 3 | Study 3: Model summary information comparing indirect effects of birthday and control conditions on stringency via psychological reactance.

	Consequent					
	M (psychological reactance)			Y (stringency in punishment)		
Antecedent	**Coefficient**	***SE***	***p***	**Coefficient**	***SE***	***p***
For all models						
M (Reactance to Essay #3)				0.42	0.19	0.028
Reactance to Essay #1	0.28	0.14	0.044	−0.68	0.34	0.048
Reactance to Essay #2	0.39	0.12	0.002	0.19	0.30	0.519
Comparing Birthday-Soliciting-Leniency to Control						
X (Birthday-Soliciting-Leniency)	0.72	0.18	<0.001	−0.11	0.45	0.813
AB (Effect of X on Y via M)				0.31	0.17	95% CI: 0.035 to 0.708
Comparing Birthday-Mentioned to Control						
X (Birthday-Mentioned)	0.34	0.18	0.057	0.09	0.42	0.835
AB (Effect of X on Y via M)				0.15	0.10	95% CI: 0.010 to 0.424
Comparing Birthday-Soliciting-Leniency to Birthday-Mentioned						
X (Birthday-Soliciting-Leniency)	0.38	0.18	0.034	−0.19	0.42	0.646
AB (Effect of X on Y via M)				0.16	0.12	95% CI: 0.004 to 0.498

N = 157 for all models. Bias-corrected confidence intervals for the indirect (AB) effects based on 5,000 bootstrap samples.

0.004 to 0.498). We ran this same set of models, including negative affect as the mediator rather than reactance, and in each case the indirect effect straddled zero, indicating that negative affect does not explain the increased stringency toward birthday offenders.

Discussion

In a substantially different paradigm and using a different type of transgression, Study 3 shows that psychological reactance to transgressors on their birthdays drives authority figures' increased stringency toward them. It is interesting to note the subtlety of the birthday manipulations in this study: even in the birthday-soliciting-leniency condition, the student did not use his birthday as an excuse for violating the rules of the essay-writing contest, but merely said they deserved the prize because it was their birthday. These subtle manipulations help strengthen our argument that merely making the social context of a birthday salient increases how stringently a transgressor will be treated by an authority figure with the discretion to do so. Additional tests confirmed that psychological reactance – the subjective perception that one's freedom is threatened – functions as a mechanism behind this effect. In contrast, we did not find empirical support for negative affect as a mechanism in this study.

Study 4: Experimental Evidence on Bias Salience as an Alternative Mechanism

A second alternative explanation that could drive our results is overcompensation for bias. When attention is drawn to factors that may bias an individual's evaluation of a target, a typical response is to try to correct for that possibility by adjusting the judgment away from the direction of the bias (Martin, 1986; Schwarz and Bless, 1992; Wegener and Petty, 1997). Given the challenges in correctly estimating the size of a potential bias, people often overcorrect for it in practice, leading to disproportional responses in the opposite direction (Wegener and Petty, 1995, 1997). We conducted another experiment in the same lab, to test whether evaluators' stringency could be explained by a concern that they might be making a biased decision when it was the transgressor's birthday.

Participants and Procedure

The behavioral lab at a UK-based business school (31% male; M_{age} = 25.7 years, SD = 8.3) recruited 101 participants to complete the study for a £10 payment. The experiment used the same experimental paradigm as Study 3, but employed a 2 (birthday-mentioned vs. control) × 2 (bias-salient vs. control) between-subjects design. The study was approved by the school's Ethics Review Board, and met all APA requirements for the ethical treatment of research participants.

This experiment used a manipulation which simply mentioned the essay-writer's birthday without actively soliciting leniency [**birthday-mentioned**: "*It would be so great to hear that I won first prize next Wednesday, because it's my birthday that day and I'm already going to be celebrating with my friends!*" vs control: "*It would be so great to hear that I won first prize next Wednesday!*"]. In the **bias-salient** condition, the participants read these additional instructions: "*The leaders at the school are concerned that bias plays a role in who the teachers normally nominate to win this tuition discount. Therefore, you, as a lab participant, are helping us to understand if this bias is occurring, and if so, how it might be affecting student outcomes. Please be aware that what we know about people can sometimes bias our assessments of them. Try to be as UNBIASED in your assessments as possible.*" If bias salience was driving our effects, we should observe an interaction between the birthday-mentioned and bias-salient conditions, such that participants were more stringent for birthday essay writers in the bias-salient condition, compared to those in the control condition.

Results

We ran a logistic regression with stringency as the dependent variable, and birthday condition, bias condition, and their interaction as independent variables. The coefficients for the bias salience condition ($B = -0.087$, expB $= 0.92$, $p = 0.93$) and the interaction of the two conditions ($B = -1.23$, expB $= 0.27$, $p = 0.31$) were not significant, indicating that making the possibility of biased evaluations more salient to the participants did not strongly affect whether they treated the writer of Essay #3 with increased stringency. However, the same logistic regression revealed a significant, negative coefficient for the birthday-soliciting-leniency condition ($B = 1.705$, expB $= 5.50$, $p = 0.046$), indicating—consistent with the findings of Study 3—that Essay Writer #3 was treated with increased stringency in the birthday condition. Together, these results are consistent with Study 3, and suggest that overcompensation for bias is not a supported alternative mechanism for our effects.

GENERAL DISCUSSION

Our evidence from the field and the lab was consistent with the predictions of our theory. When confronted with a social expectation of lenient treatment, individuals with the authority and discretion to punish them treat transgressors more stringently rather than more leniently. In our studies, we used the transgressor's birthday as a social context that leads to an expectation of lenient treatment, as it has many attractive characteristics that allow us to test this phenomenon in the field. Increased stringency for birthday transgressors, which we identified both in the field and in the lab, runs counter to what individuals believe happens to transgressors on their birthdays. We find that this effect is driven by psychological reactance toward the transgressor. Moreover, psychological reactance increases as the salience of the transgressor's birthday increases (as, we assume, is the perception that the target is using his birthday to actively solicit lenient treatment).

Theoretical Implications

Our results contribute to several literatures. First, our research contributes to a broader literature on punishment from psychological perspectives (Treviño, 1992; Fragale et al., 2009). Though theory has offered frameworks to evaluate when and how to punish (Arvey and Jones, 1985; Butterfield et al., 1996), examined its consequences (Ball et al., 1994; Podsakoff et al., 2006), and looked at how aspects of organizational context shape punishment decisions (Beyer and Trice, 1984), we know less about how the social context of transgressions affects punishment decisions. Even the literature on just deserts, which focuses on motivations to punish, has only addressed aspects of the social context that directly speak to the harm the act has caused or justifiable mitigating circumstances for it (such as the difference between intentional and accidental actions). However, there are many aspects of our context that may affect motivations to punish and punishment decisions, with only tenuous relevance to the transgression. We know little about aspects of our social context that ought to be unrelated to punishment decisions affect those decisions nevertheless. Our research addresses this gap by showing how subtle contextual factors (it being the transgressor's birthday) play an important role in the ultimate penalties authorities impose.

Second, this paper contributes to our knowledge of how individuals behave when a context elicits two different motivations with conflicting behavior prescriptions. The large body of work in both psychology (e.g., Cialdini et al., 1991) and economics (e.g., Fehr and Gächter, 2002) that examines the power of expectations on individual behavior has focused primarily on how a single social norm motivates behavior. Instead, our work examines how individuals respond to multiple expectations elicited by the social context and that motivate us in conflicting ways, and how these motivational conflicts they may influence behavior.

Third, these findings extend our understanding of psychological reactance among individuals with the discretion over penalizing transgressions. Most of the literature on reactance has focused on refusals to help others (Berkowitz, 1973), engage in more positive behaviors, such as healthier lifestyle choices (Dillard and Shen, 2005), or pursue goals (Chartrand et al., 2007). These findings show that reactance also drives behaviors in punishment contexts, and confirms again that even very subtle messages can elicit perceptions that one's freedom is being threatened, driving our behavior in the opposite direction.

Ultimately, these findings help us understand how discretion is exercised in the field, thus deepening our knowledge of how discrimination operates. Work on discrimination has focused almost exclusively on demographic characteristics such as age, race, ethnicity, and gender (Paluck and Green, 2009). Our research shows that other, less obvious factors will also lead individuals to treat transgressions differentially. This suggests that we need to extend our vigilance about how discretion may undermine the efficiency of punishment. It also deepens our understanding about the challenges humans have in debiasing their judgments and behavior (Wegener and Petty, 1995, 1997), particularly when individuals with discretion over how someone is treated interact with that person in advance of imposing penalties on them.

Practical Implications

Our research also has important practical implications, both for alleged transgressors as well as those with discretion over punishing them—from managers and teachers to judges and jury members. Transgressors need to be aware that their intuitions about avoiding punishment by making leniency norms salient may backfire, resulting in harsher penalties than if they refrained from making the norm salient. In other words, transgressors may benefit from avoiding any perception that they are trying to capitalize on contextual factors that would suggest more lenient treatment. On the other hand, authorities with the responsibility to punish should be aware that in the face of conflicting motivations, they may make decisions that undermine the fairness, and ultimately the effectiveness, of their sanctions.

Compared to others responsible for punishing transgressions, law enforcement officials may be particularly likely to react negatively to perceptions that offenders are soliciting lenient treatment. They are accustomed to excuses and pleas for leniency from those they penalizing, to the extent that such pleas can become tiresome and prompt cynicism (Van Maanen, 1974). Research on leniency in law enforcement suggests that officer have "pet peeves," including many related to the demeanor of offenders, that may elicit reactive and more severe responses (Schafer and Mastrofski, 2005). Contrition and verbally accepting responsibility for one's actions may elicit more lenient responses from law enforcement, while soliciting special treatment may trigger reactive responses (Schafer and Mastrofski, 2005). Indeed, recent work by van Prooijen and Kerpershoek (2011) suggests that individuals may inflict excess retribution when given discretion to punish criminals, but only when they feel their autonomy is threatened. Though our data do not allow us to observe what specifically happens in the dyadic interactions between drivers and officers in our field data, our scenario studies suggest that driver behavior, and subsequent officer reactions to those behaviors, are critical to outcomes.

Our research also has important practical implications for managers, who are commonly given broad discretion to punish employees through oral reprimands, work suspension, or, in extreme cases, termination (Beyer and Trice, 1984; Butterfield et al., 1996). In fact, punishment is a widely used managerial strategy for producing desired changes in employee behavior (Ball et al., 1994). Thus, it is important for managers to know that they are also vulnerable to the challenges associated with managing contradictory motivations that might influence their actions.

Conclusion

Our findings suggest that when authority figures have discretion over punishment decisions, making an expectation that a transgressor will be treated leniently salient leads to a negative psychological reaction, leading individuals with the authority to punish to do so more harshly. This might lead to the conclusion that discretion is overrated or overused. Yet, we do not want to suggest that our findings provide an argument against discretion, only a fair warning about some of its additional problematic qualities. Many dysfunctional consequences result when discretion is unavailable, such as under mandatory punishment guidelines (e.g., "three strikes" laws). These consequences include higher rates of violence and murders among repeat offenders and against witnesses of repeat offenses (Marvell and Moody, 2001; Zimring et al., 2001). Thus, eliminating discretion is likely not the answer. The message we take from our findings is that authorities with discretion over punishment should be vigilant about how the situational cues may be affecting their psychological reactions to the transgressors and ultimately, their punishment decisions.

AUTHOR CONTRIBUTIONS

The authors contributed equally to this work. LP collected, analyzed, and wrote up the field data. LP and CM conceived the experiments. CM ran the experiments, analyzed and wrote up the lab data. LP and CM discussed the results and implications and commented on the manuscript at all stages, and wrote the paper jointly.

ACKNOWLEDGMENTS

We thank Nicholas P. Lovrich and the Division of Government Studies and Services at Washington State University for the opportunity to work with the Washington State Patrol on this project. Brady Horn and Dick Doane provided considerable help in data provision. We also thank Francesca Gino, and participants at the 2012 Academy of Management Meeting in Boston as well as the Behavioral Ethics Conference at the University of Central Florida for their insightful comments on presentations based on earlier drafts of this paper. Finally, we thank the Edmond J. Safra Center for Ethics, London Business School and the Olin Business School at the Washington University in St. Louis for their financial support of this research.

REFERENCES

Ammons, L. L. (1994). Discretionary justice: a legal and policy analysis of a governor's use of the clemency power in the cases of incarcerated battered women. *J. Law Policy* 3:1.

Arvey, R. D., and Jones, A. P. (1985). The use of discipline in organizational settings: a framework for future research. *Res. Organ. Behav.* 7, 367–408.

Ball, G. A., Treviño, L. K., and Sims, H. P. (1994). Just and unjust punishment: influences on subordinate performance and citizenship. *Acad. Manag. J.* 37, 299–322. doi: 10.2307/256831

Berkowitz, L. (1969). Resistance to improper dependency relationships. *J. Exp. Soc. Psychol.* 5, 283–294. doi: 10.1016/0022-1031(69)90054-7

Berkowitz, L. (1973). Reactance and the unwillingness to help others. *Psychol. Bull.* 79, 310–317. doi: 10.1037/h0034443

Beyer, J. M., and Trice, H. M. (1984). A field study of the use and perceived effects of discipline in controlling work performance. *Acad. Manag. J.* 27, 743–764. doi: 10.2307/255876

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. New York: Academic Press.

Brehm, J. W., and Cole, A. H. (1966). Effect of a favor which reduces freedom. *J. Pers. Soc. Psychol.* 3, 420–426. doi: 10.1037/h0023034

Brehm, S. S., and Brehm, J. W. (1981). *Psychological Reactance: A Theory of Freedom and Control*. New York, NY: Academic Press.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* 6, 3–5. doi: 10.1177/1745691610393980

Butterfield, K. D., Treviño, L. K., and Ball, G. A. (1996). Punishment from the manager's perspective: a grounded investigation and inductive model. *Acad. Manag. J.* 39, 1479–1512. doi: 10.2307/257066

Carlsmith, K. M., and Darley, J. M. (2008). *Psychological Aspects of Retributive Justice Advances in Experimental Social Psychology*, Vol. 40. Cambridge: Academic Press, 193–236.

Carlsmith, K. M., Darley, J. M., and Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *J. Pers. Soc. Psychol.* 83, 284–299. doi: 10.1037/0022-3514.83.2.284

Chartrand, T. L., Dalton, A. N., and Fitzsimons, G. J. (2007). Nonconscious relationship reactance: when significant others prime opposing goals. *J. Exp. Soc. Psychol.* 43, 719–726. doi: 10.1016/j.jesp.2006.08.003

Cialdini, R. B. (2009). *Influence: Science and Practice*, 5th Edn. Boston: Pearson Education.

Cialdini, R. B., Kallgren, C. A., and Reno, R. R. (1991). A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. *Adv. Exp. Soc. Psychol.* 24, 201–234. doi: 10.1016/S0065-2601(08)60330-5

Dillard, J. P., and Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Commun. Monogr.* 72, 144–168. doi: 10.1080/03637750500111815

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

Fitzsimons, G. J., and Lehmann, D. R. (2004). Reactance to recommendations: when unsolicited advice yields contrary responses. *Market. Sci.* 23, 82–94. doi: 10.1287/mksc.1030.0033

Fragale, A. R., Rosen, B., Xu, C., and Merideth, I. (2009). The higher they are, the harder they fall: the effects of wrongdoer status on observer punishment recommendations and intentionality attributions. *Organ. Behav. Hum. Decis. Process* 108, 53–65. doi: 10.1016/j.obhdp.2008.05.002

Gibbons, F. X., and Wicklund, R. A. (1982). Self-focused attention and helping behavior. *J. Pers. Soc. Psychol.* 43, 462–474. doi: 10.1007/s10508-008-9370-9

Goffman, E. (1955). On face-work: an analysis of ritual elements in social interaction. *Psychiatry* 18, 213–231.

Goodman, J. K., Cryder, C. E., and Cheema, A. (2013). Data collection in a flat world: the strengths and weaknesses of mechanical turk samples. *J. Behav. Decis. Mak.* 26, 213–224. doi: 10.1002/bdm.1753

Greene, D., Barber, L., Chorney, M., Martyn, B., Tanney, A., Thurston, N., et al. (1987). Birthdays. *J. Psychosoc. Nurs. Ment. Health Serv.* 25, 8–9.

Hauser, D. J., and Schwarz, N. (2015). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behav. Res. Methods* 48, 400–407. doi: 10.3758/s13428-015-0578-z

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford.

Horn, B. P., McCluskey, J. J., and Mittelhammer, R. C. (2014). Quantifying bias in driving-under-the-influence enforcement. *Econ. Inq.* 52, 269–284. doi: 10.1111/ecin.12043

Imbens, G., and Lemieux, T. (2008). Regression discontinuity designs: a guide to practice. *J. Econ.* 142, 615–635. doi: 10.1016/j.jclinepi.2014.06.021

Jiobu, R. M., and Knowles, E. S. (1974). Norm strength and alms giving: an observational study. *J. Soc. Psychol.* 94, 205–211. doi: 10.1080/00224545.1974.9923207

Kadish, S. H. (1961). Legal norm and discretion in the police and sentencing processes. *Harv. Law Rev.* 75, 904–931. doi: 10.2307/1338645

Karelaia, N., and Keck, S. (2012). *When Deviant Leaders are Punished More Than Non-Leaders: The Role of Deviance Severity*. INSEAD Working Paper No. 2012/120/DS. Available at: http://ssrn.com/abstract=1747642

Keltner, D., and Anderson, C. (2000). Saving face for darwin: the functions and uses of embarrassment. *Curr. Dir. Psychol. Sci.* 9, 187–192. doi: 10.1111/1467-8721.00091

LaFave, W. R. (1970). The prosecutor's discretion in the United States. *Am. J. Comp. Law* 18, 532–548. doi: 10.2307/839344

Langer, E. J., Blank, A., and Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: the role of 'placebic' information in interpersonal interaction. *J. Pers. Soc. Psychol.* 36, 635–642. doi: 10.1037/0022-3514.36.6.635

Levitt, S. D., and Porter, J. (2001). How dangerous are drinking drivers? *J. Polit. Econ.* 109, 1198–1237. doi: 10.1086/323281

Lovrich, N. P., Gaffney, M. J., Mosher, C. C., Pratt, T. C., and Pickerill, M. J. (2007). *Results of the Monitoring of WSP Traffic Tops for Biased Policing*. Washington, DC: Division of Governmental Studies and Services, Washington State University. Available at: http://www.wsp.wa.gov/publications/reports/wsu_2007_report.pdf

Martin, L. L. (1986). Set/reset: use and disuse of concepts in impression formation. *J. Pers. Soc. Psychol.* 51, 493–504. doi: 10.1037/0022-3514.51.3.493

Marvell, T. B., and Moody, C. E. (2001). The lethal effects of three-strikes laws. *J. Legal Stud.* 30, 89–106. doi: 10.1086/468112

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York, NY: Harper & Row.

Miron, A. M., and Brehm, J. W. (2006). Reactance Theory - 40 Years Later. *Z. Sozialpsychol.* 37, 9–18. doi: 10.1024/0044-3514.37.1.9

National Highway Traffic Safety Administration (2010). *Traffic Safety Facts 2010: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. Washington, DC: U.S. Government Printing Office.

Paluck, E. L., and Green, D. P. (2009). Prejudice reduction: what works? A review and assessment of research and practice. *Annu. Rev. Psychol.* 60, 339–367. doi: 10.1146/annurev.psych.60.110707.163607

Piehl, J. (1977). Integration of information in the "court:" influence of physical attractiveness on amount of punishment for a traffic offender. *Psychol. Rep.* 41, 551–556. doi: 10.2466/pr0.1977.41.2.551

Pierce, L., Dahl, M. S., and Nielsen, J. (2013). In sickness and in wealth: psychological and sexual costs of income comparison in marriage. *Pers. Soc. Psychol. Bull.* 39, 359–374. doi: 10.1177/0146167212475321

Pierce, L., and Snyder, J. A. (2012). Discretion and manipulation by experts: evidence from a vehicle emissions policy change. *B.E. J. Econ. Anal. Policy* 13, 1–30. doi: 10.2139/ssrn.1831494

Podsakoff, P. M., Bommer, W. H., Podsakoff, N. P., and MacKenzie, S. B. (2006). Relationships between leader reward and punishment behavior and subordinate attitudes, perceptions, and behaviors: a meta-analytic review of existing and new research. *Organ. Behav. Hum. Decis. Process* 99, 113–142. doi: 10.1016/j.obhdp.2005.09.002

Police Executive Research Forum (2001). *Racially Biased Policing—A Principled Response*. Washington, DC: Justice Department's Office of Community Oriented Policing. Available at: http://www.cops.usdoj.gov/Publications/RacialyBiasedPolicing.pdf

Reiss, A. J. (1984). Consequences of compliance and deterrence models of law enforcement for the exercise of police discretion. *Law Contemp. Probl.* 47, 83–122. doi: 10.2307/1191688

Reno, R. R., Cialdini, R. B., and Kallgren, C. A. (1993). The transsituational influence of social norms. *J. Pers. Soc. Psychol.* 64, 104–112. doi: 10.1037/0022-3514.64.1.104

Robinson, P. H., and Darley, J. M. (1997). The utility of desert. *Northwest Univ. Law Rev.* 91, 453–499.

Schafer, J. A., and Mastrofski, S. D. (2005). Police leniency in traffic enforcement encounters: exploratory findings from observations and interviews. *J. Crim. Justice* 33, 225–238. doi: 10.1016/j.jcrimjus.2005.02.003

Schwarz, N., and Bless, H. (1992). "Constructing reality and its alternatives: an inclusion/exclusion model of assimilation and contrast effects in social judgment," in *The Construction of Social Judgments*, eds L. L. Martin and A. Tresser (Hillsdale, NJ: Lawrence Erlbaum), 217–245.

Sherman, L. (1984). Experiments in police discretion: scientific boon or dangerous knowledge? *Law Contemp. Probl.* 47, 61–81. doi: 10.2307/1191687

Sigall, H., and Ostrove, N. (1975). Beautiful but dangerous: effects of offender attractiveness and nature of the crime on juridic judgment. *J. Pers. Soc. Psychol.* 31, 410–414. doi: 10.1037/h0076472

Smith, M. R., and Alpert, G. P. (2007). Explaining police bias. *Crim Justice Behav.* 34, 1262–1283. doi: 10.1177/0093854807304484

Snyder, C. R., Stucky, R. J., and Higgins, R. L. (1983). *Excuses: Masquerades in Search of Grace*. New York, NY: Wiley.

Snyder, J. (2010). Gaming the liver transplant market. *J. Law Econ. Organ.* 26, 546–568. doi: 10.1093/jleo/ewq003

Steinberg, L., and Scott, E. S. (2003). Less guilty by reason of adolescence: developmental immaturity, diminished responsibility, and the juvenile death penalty. *Am. Psychol.* 58, 1009–1018. doi: 10.1037/0003-066X.58.12.1009

Stewart, J. E. (1985). Appearance and punishment: the attraction-leniency effect in the courtroom. *J. Soc. Psychol.* 125, 373–378. doi: 10.1080/00224545.1985.9922900

Treviño, L. K. (1992). The social effects of punishment in organizations: a justice perspective. *Acad. Manag. Rev.* 17, 647–676. doi: 10.5465/AMR.1992.4279054

Van Maanen, J. (1974). "Working the street: a developmental view of police behavior," in *Control in the Police Organization*, ed. H. Jacob (Cambridge, MA: MIT Press), 275–317.

van Prooijen, J.-W., and Kerpershoek, E. F. P. (2011). The impact of choice on retributive reactions: how observers' autonomy concerns shape responses to criminal offenders. *Br. J. Soc. Psychol.* 52, 329–344. doi: 10.1111/j.2044-8309.2011.02079.x

Vorenberg, J. (1976). Narrowing the discretion of criminal justice officials. *Duke Law J.* 1976, 651–697. doi: 10.2307/1371934

Waldfogel, J. (1993). The deadweight loss of Christmas. *Am. Econ. Rev.* 83, 1328–1336.

Wegener, D. T., and Petty, R. E. (1995). Flexible correction processes in social judgment: the role of naive theories in corrections for perceived bias. *J. Pers. Soc. Psychol.* 68, 36–51. doi: 10.1037/0022-3514.68.1.36

Wegener, D. T., and Petty, R. E. (1997). The flexible correction model: the role of naive theories of bias in bias correction. *Adv. Exp. Soc. Psychol.* 29, 141–208. doi: 10.1016/S0065-2601(08)60017-9

Zimring, F. E., Hawkins, G., and Kamin, S. (2001). *Punishment and Democracy: Three Strikes and You're out in California*. New York, NY: Oxford University Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 April 2015
doi: 10.3389/fpsyg.2015.00535

Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior

*Camilla J. Rajah-Kanagasabai and Lynne D. Roberts**

School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia

This study examined the utility of the Theory of Planned Behavior model, augmented by descriptive norms and justifications, for predicting self-reported research misconduct and questionable research practices in university students. A convenience sample of 205 research active Western Australian university students (47 male, 158 female, ages 18–53 years, $M = 22$, SD = 4.78) completed an online survey. There was a low level of engagement in research misconduct, with approximately one in seven students reporting data fabrication and one in eight data falsification. Path analysis and model testing in LISREL supported a parsimonious two step mediation model, providing good fit to the data. After controlling for social desirability, the effect of attitudes, subjective norms, descriptive norms and perceived behavioral control on student engagement in research misconduct and questionable research practices was mediated by justifications and then intention. This revised augmented model accounted for a substantial 40.8% of the variance in student engagement in research misconduct and questionable research practices, demonstrating its predictive utility. The model can be used to target interventions aimed at reducing student engagement in research misconduct and questionable research practices.

Keywords: research misconduct, data fabrication, data falsification, academic integrity, Theory of Planned Behavior, descriptive norms, justifications, questionable research practices

OPEN ACCESS

Edited by:
Guy Hochman,
Duke University, USA

Reviewed by:
Tom Stone,
Oklahoma State University, USA
Eyal Peer,
Bar-Ilan University, Israel

****Correspondence:***
Lynne D. Roberts,
School of Psychology and Speech Pathology, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
lynne.roberts@curtin.edu.au

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *04 November 2014*
Accepted: *14 April 2015*
Published: *30 April 2015*

Citation:
Rajah-Kanagasabai CJ and Roberts LD (2015) Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. Front. Psychol. 6:535. doi: 10.3389/fpsyg.2015.00535

Introduction

Academic integrity is vital to the foundation of the academic community and its credibility (McCabe and Trevino, 1993; McCabe et al., 2008). There are two types of dishonest misconduct that threaten academic integrity: academic misconduct (cheating, deception, and corruption; Mavrinac et al., 2010) and research misconduct (fabrication, falsification, and plagiarism in proposing and conducting research or reporting results; National Health and Medical Research Council and Australian Research Council, 2007). The US Department of Health and Human Services Office of Research Integrity (2000, p. 1) further define fabrication as making up data or results and reporting them, and falsification as "manipulating research materials, processes or changing or omitting data." Questionable research practices, consisting of failing to obtain approval, not obtaining consent before conducting research, ignoring outliers, publishing *post hoc* analyses without explanation, and publishing articles using data that have not been collected legitimately or that

have been reported elsewhere (Pimple, 2002; Gilbert and Denison, 2003; Martinson et al., 2005; Kumar, 2008; Rose, 2008; Bedian et al., 2010), also fall within the umbrella of research misconduct. While data fabrication and falsification are the more serious forms of research misconduct, questionable research practices potentially have a larger impact on research integrity as they are more widespread (Anderson et al., 2013).

A growing body of research has examined research misconduct in academic settings. The most common form of research misconduct, plagiarism, is the area of research misconduct that has received the most attention (e.g., Park, 2003; Bennett, 2005; Marsden et al., 2005; Pickard, 2006; Mavrinac et al., 2010; Ogilvie and Stewart, 2010). In comparison, limited research has addressed fabrication, falsification, and questionable research practices in academic settings, and these areas are the focus of this research.

Estimates of the prevalence of research misconduct and questionable research practices among researchers and academics range widely, depending upon the measure used. Only 20–30 cases are reported to the US National Science Foundation and Department of Health and Human Service each year, representing a rate of 1 case per 100,000 researchers (Steneck, 2006). Estimates based on journal articles retracted for fabrication or falsification provide higher prevalence rates, but vary according to the years and databases covered. Based on analysis of article retractions in journals indexed by PubMed, Claxton (2005) estimated research misconduct was detected in less than one case per 5,000 papers (0.02%). Working on the assumption that for every case detected up to 10 cases may go undetected, Claxton estimated that the actual rate of fraudulent papers may be as high as 0.2%. Across databases, Grieneisen and Zhang (2012) identified 4449 articles retracted between 1928 and 2011, reporting that 20% were retracted for research misconduct, with a further 42% retracted for questionable data or interpretation. In contrast, using only articles indexed in PubMed, Fang et al. (2012) reported that 43% of the 2,047 articles retracted were retracted for fraud or suspected fraud. Articles retracted for data fabrication and/or falsification, in comparison to articles retracted for error, are clustered in high impact journals, have more authors and the first author is more likely to have previous retractions (Steen, 2010). Across retraction studies, the incidence of retracted papers is consistently reported to be increasing over time (Steen, 2011; Fang et al., 2012; Grieneisen and Zhang, 2012).

Higher prevalence estimates again are obtained when using self-report methodologies. In a recent meta-analysis, Fanelli (2009) reported that ∼2% of scientists admitted to fabrication, falsification or modification of data at least once, whereas approximately a third admitted to questionable research practices. Interestingly, participants reported higher rates of awareness of at least one other researcher engaging in the fabrication of data (14%) and questionable research practices (72%). Further, self-reports may underestimate the actual prevalence of research misconduct and questionable research practices. John et al. (2012) provided incentives for honest reporting combined with anonymous reporting, with US academic psychologist respondents self-admitted questionable research practices ranging from 4.5% (claiming results unaffected by demographic variables when unsure/know false) to 66.5% (failing to report all of a study's dependent variables).

Research misconduct and questionable research practices by researchers and academics may have roots in practices developed while students, and may reach back as far as the undergraduate years. Studies that have explored fabrication, falsification or questionable research practices in student populations have generally used student samples from degrees in 'hard sciences,' such as biomedical science, where the 'correct' answers to laboratory experiments are already known, making results more likely to be falsified (Davidson et al., 2001). Davidson et al. (2001) reported that 40–75% of undergraduate students admitted to 'almost always' manipulating data in science labs. Similar figures have been reported for other samples of science undergraduates (Franklyn-Stokes and Newstead, 1995; Lawson et al., 1999/2000). In contrast, figures are much lower (approximately one in five) when sampling undergraduates more broadly across disciplines outside of the sciences (Brimble and Stevenson-Clarke, 2005; McCabe, 2005). Of particular concern, one in ten Ph.D. students report falsification and fabrication of data is acceptable (Hofman et al., 2013).

Students who engage in academically dishonest behavior at university are likely to engage in dishonest behavior in the workforce (Nonis and Swift, 2001; Graves, 2008), highlighting the importance of understanding and addressing research misconduct at the time it first emerges, in the undergraduate years.

In attempting to understand dishonest behavior a range of competing economic, criminological and psychological theories have been used. In summarizing the factors shaping dishonest behavior across contexts, Ariely (2012, Figure 6) highlights the role of rationalizations, conflicts of interest, creativity, engaging in the first dishonest act, ego-depletion, benefit to others, observing the dishonest behavior of others and culture. Within academic settings, a range of theoretical frameworks, such as the General Theory of Crime (Gottfredson and Hirshci, 1990), Social Learning Theory (Bandura, 1978), Techniques of Neutralization (Sykes and Matza, 1957), Multidimensional Ethics Theory (Yang, 2012b) and the Theory of Planned Behavior (Ajzen, 1985) have been successfully applied in understanding academic dishonesty, but little research has focused on predicting fabrication, falsification and questionable research practices in university students. Of these theories, the Theory of Planned Behavior has consistently had good explanatory power, explaining 33–48% of the variance in health, social, and economic behavior (Armitage and Conner, 2001) and may be usefully applied to predicting engagement in research misconduct and questionable research practices.

Theory of Planned Behavior

The Theory of Planned Behavior posits that intention drives behavior, with attitudes toward the behavior and subjective norms influencing behavior through intention, and perceived behavioral control impacting behavior both directly and mediated through intention (Ajzen, 1991). Attitudes represent positive or negative beliefs about behavior and its consequences. If a behavior is judged positively, attitude increases intention to

engage in that behavior. Subjective norms represent perceived pressure from others to engage in behavior, and increase intention to engage in the behavior. Perceived behavioral control represents the perceived difficulty in performing the behavior, with greater difficulty reducing both intention to engage in behavior and actual behavior. Attitudes, subjective norms and perceived behavioral control form intention to perform a behavior, which if strong enough, will result in engagement (Ajzen, 1991). Ideally behavior is measured at a later point in time than intention, however, previous research has indicated that past behavior can be used as a proxy for future behavior (Rise et al., 2010).

Whilst not previously used to predict engagement in research misconduct and questionable research practices, the Theory of Planned Behavior has been used to predict cheating by undergraduate students. An early study by Beck and Ajzen (1991) used the Theory of Planned Behavior to predict a range of dishonest actions, including cheating on a test or exam. The Theory of Planned Behavior explained 67% of the variance in cheating intention and 55% of the variance in cheating behavior. However, subjective norms was not a significant predictor of intention and perceived behavioral control was not a significant predictor of behavior. Stone et al. (2009, 2010) examined cheating by undergraduate business students. The Theory of Planned Behavior explained 21% and 36% of the variance in cheating intention and cheating behavior respectively (Stone et al., 2010). Alleyne and Phillips (2011) examined undergraduate students' intention to cheat and lie, reporting that Theory of Planned Behavior variables accounted for 48% of intention to cheat and 29% of intention to lie (actual behavior was not measured). Harding et al. (2007) found general support for the Theory of Planned Behavior model in predicting undergraduate cheating, but perceived behavioral control was not a significant predictor of behavior. In a further study, Mayhew et al. (2009) reported that neither attitudes nor perceived behavioral control were significant predictors of intention or behavior when moral obligation was added to the Theory of Planned Behavior model.

Extending the Theory of Planned Behavior Model

A major strength of the Theory of Planned Behavior is that variables can be added to the model to increase its explanatory power (Ajzen, 1985). Two variables of interest in predicting engagement in research misconduct and questionable research practices are descriptive norms and justifications.

Descriptive norms relate to what others actually do (Rivis and Sheeran, 2003). As such, they represent the individual's perception of behavior by others, in contrast to the traditional injunctive conceptualization of subjective norms where the focus is on the individual's perception of perceived pressure from others to engage in a particular behavior (Ajzen, 1991). The distinction has been described in terms of 'what is' (descriptive norms) versus 'what ought' (subjective norms; also known as injunctive norms, Cialdini et al., 1990) to be done (Forward, 2009). Behavior is influenced by whether injunctive or descriptive norms are salient within a particular setting (Cialdini et al., 1990; Kallgren et al., 2000). Behavior by in-group members invokes descriptive norms, while behavior by out-group members invokes injunctive norms (Gino et al., 2009). Behavior is also influenced by the extent to which actions violate the salient norm and the personal norms of the individual (Kallgren et al., 2000). While injunctive norms may influence behavior across settings, descriptive norms influence behavior only in settings where they are salient (Reno et al., 1993). In more recent reconceptualizations of the structure of the Theory of Planned Behavior predictor variables, Fishbein and colleagues (Fishbein, 2000; Fishbein and Yzer, 2003; Ajzen and Fishbein, 2005) have noted the need to include both injunctive and descriptive norms "in order to obtain a complete measure of subjective norm" (Ajzen and Fishbein, 2005, p. 199). However, this practice does not appear to have been routinely adopted, with some research indicating injunctive and subjective norms are conceptually distinct and differentially predict intention and behavior (Forward, 2009; Manning, 2009).

Meta-analytic findings provide further support for the addition of descriptive norms to the Theory of Planned Behavior model. Descriptive norms and intention are medium-to-strongly correlated ($r = 0.44$) and account for an additional 5% of the variance in intention across a range of behaviors, after controlling for attitudes, subjective norms and perceived behavioral control (Rivis and Sheeran, 2003). However, descriptive norms were not predictive of intention for all behaviors, with moderator analyses indicating descriptive norms are of most importance in predicting intention to engage in risk behaviors and with younger samples (Rivis and Sheeran, 2003). Research predicting student engagement in research misconduct and questionable research practices meets both these criteria. A further meta-analysis by Manning (2009) indicated that the relationship between descriptive norms and behavior is stronger than the relationship between subjective norms and behavior, and that in modeling the Theory of Planned Behavior there is a direct path from descriptive norms to behavior, but only a mediated path from subjective norms to behavior. Descriptive norms have previously been demonstrated to be significantly correlated with both intention to engage in academic misconduct ($r = 0.37$) and actual academic misconduct ($r = 0.49$; Stone et al., 2010[1]), further justifying their addition to the Theory of Planned Behavior model.

As behaviors such as engaging in academic and research misconduct are not based on honest errors of judgment, individuals need to justify their engagement in the behavior (Stone et al., 2009). The mismatch between beliefs and behavior creates cognitive dissonance (Festinger, 1957), a psychological state that creates discomfort to the individual and motivates change to reduce the dissonance. More specifically, the term 'ethical dissonance' is used to describe cognitive dissonance resulting from behaviors deviating from accepted social norms (Barkan et al., 2012; Shalvi et al., 2015). Dissonance can be resolved through changing beliefs, changing behavior, adding new attitudes consistent with the behavior, or devaluing the importance of the dissonance (Festinger, 1957). Justifications may act to reduce

[1]Stone et al. (2010) labeled their normative measure 'Subjective norm,' but the items all reflect descriptive norms.

dissonance through devaluating the importance of the dissonance (Stone et al., 2009). Self-serving justifications may reduce ethical dissonance through redefining and excusing questionable behaviors prior to engagement, or through compensatory mechanisms following engagement. Whether pre- or post-behavior, justifications attenuate the threat to the moral self (Shalvi et al., 2015).

Possible justifications for engaging in academic misconduct and questionable research practices include perceptions of others engaging in academic misconduct, helping a friend, peer pressure, extenuating circumstances and fear of failure (Stone et al., 2009). Stone et al. (2009) argue that justifications are used by those who have already engaged in academic misconduct, and play a potentially mediating role between the Theory of Planned Behavior predictor variables of attitudes, subjective norms and perceived behavioral control and the outcome variable of academic misconduct. In their study examining students' cheating behavior, Stone et al. (2009) reported that attitudes, subjective norms and perceived behavioral control accounted for 28% of the variance in justifications, which in turn was a significant predictor of cheating behavior. Justifications were strongly correlated with both intention ($r = 0.60$) and behavior ($r = 0.54$). As academic and research misconduct are related constructs, this study provides strong support for the augmentation of the Theory of Planned Behavior model with justifications in predicting student engagement in research misconduct and questionable research practices.

Demographic factors may also be important in understanding student engagement in research misconduct and questionable research practices. Factors that have been explored in relation to this type of dishonest behavior are age, gender, and year of study. Negative correlations between age and academic misconduct have been reported (Brimble and Stevenson-Clarke, 2005), but inconsistent results found in relation to gender (Davidson et al., 2001; Yang, 2012a). A higher prevalence of research misconduct has been observed in lower year students (Yang, 2012b). Additionally, social desirability is an important construct to measure in self-report studies exploring research misconduct (Jann et al., 2012) as research misconduct is widely considered to be an unethical practice (Arvidson, 2004) and may elicit socially desirable responses.

In summary, there is limited research examining the predictors of student engagement in research misconduct and questionable research practices. The Theory of Planned Behavior is one model that may have utility in understanding these behaviors. Previous research that has examined the Theory of Planned Behavior in relation to academic integrity has mainly focused on cheating, but has demonstrated good explanatory power in some studies (Stone et al., 2009; Alleyne and Phillips, 2011). Drawing together previous disparate research on predictors of dishonest behavior into an integrated model applied to academic integrity, this study will examine the predictive utility of the Theory of Planned Behavior model augmented by descriptive norms and justifications (see **Figure 1**) in describing student engagement in research misconduct and questionable research practices. It is hypothesized that after controlling for demographic variables (age, gender, years of study) and social desirability, intention and justification will mediate the relationships between attitudes, subjective norms and descriptive norms with behavior (engaging in research misconduct and questionable research practices), and partially mediate the relationship between perceived behavioral control with behavior.

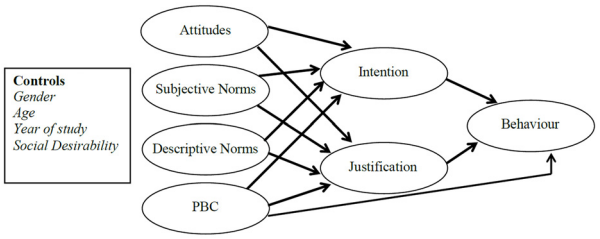

FIGURE 1 | Model of theory of planned behavior augmented by descriptive norms and justifications.

Materials and Methods

Research Design

This study used a self-report, correlational design to examine whether intention and justification (mediator variables) mediate the relationship between attitudes, subjective norms, descriptive norms, and perceived behavioral control (predictor variables) and student engagement in research misconduct and questionable research practices (criterion variable) while controlling for age, gender, years of study, and social desirability.

Participants

A non-probability, convenience sample of Western Australian university students aged 18 years and older who had collected data or conducted research for an assignment or dissertation were recruited. The final sample consisted of 205 participants from five Western Australian universities (47 male, 158 female), aged between 18 and 53 years ($M = 22$, SD $= 4.78$). The majority of students sampled had a major or minor in Psychology (71.7%) and were from one university (84.8%). Years of completed study in university ranged from half a year to 9 years ($M = 2.54$, SD $= 1.46$). An a-priori power analysis (power 0.80, alpha 0.05) indicated that based on partial correlations of previous analyses (Stone et al., 2009), a sample size of 200 participants would be required to detect a 'moderate' mediation effect (Soper, 2013). The sample obtained exceeded this estimate and was deemed sufficient for testing mediation (Tabachnick and Fidell, 2007).

Measures

An online questionnaire consisting of eight scales was developed using Qualtrics software. **Table 1** provides a summary of the measures, number of items, example items, response formats and Cronbach's alpha for each measure. At the beginning of the survey, and at the top of most pages of the survey, the following definition of research misconduct was provided:

TABLE 1 | Details of scale measures ($N = 205$).

Variable	Scale	No. of Items	Example Item (How responses were measured)	Scale range	α	Mean (SD)
Behavior	Adapted from Yang (2012a)	9[a]	How many times have you falsified results? (four point frequency scale – 1 = *never*, 2 = *one or two times*, 3 = *three to five times* and 4 = *six or more times*)	1–3	0.91	1.15 (0.29)
Attitudes	Adapted from Stone et al. (2009)	6[b]	It is always wrong to engage in research misconduct (five-point Likert scale – 1 = *strongly disagree* and 5 = *strongly agree)*	1–4	0.81	2.17 (0.63)
Subjective norms	Adapted from Beck and Ajzen (1991)	3	If I engaged in research misconduct, most people who are important to me would" (7-point Likert scale – 1 = *not care* and 7 = *disapprove*)	1–7	0.74	5.33 (1.47)
Descriptive norms	Adapted from Stone et al. (2009)	4[c]	**Quantity item** – Approximately what percentage of students do you think engage in some kind of research misconduct? (open response) **Frequency item** – How frequently do you think research misconduct occurs in classes at your university? (1 = *never*, 2 = *less than once a month*, 3 = *once a month*, 4 = *2–3 times a month*, 5 = *once a week*, 6 = *2–3 times a week* and 7 = *daily*)	0–100		26.46 (20.65)
Perceived behavioral control	Adapted from Stone et al. (2009)	4	It is easy to engage in research misconduct and not get caught (5-point Likert scale – 1 = *strongly disagree* and 5 = *strongly agree*)	1–5	0.89	2.83 (0.97)
Intention	Adapted from Yang (2012a)	9	How likely are you in the next year, to falsify results (5-point Likert scale – 1 = *very unlikely* and 5 = *very likely)*	1–4	0.91	1.51 (0.63)
Justifications	Adapted from Stone et al. (2009)	9	How likely are you to engage in research misconduct, because of laziness (5-point Likert scale – 1 = *very unlikely* and 5 = *very likely)*	1–4	0.92	1.96 (0.79)
Social Desirability	Adapted from Francis et al. (1992)	12[d]	Do you always practice what you preach? (dichotomous scale – 1 = *no* and 2 = *yes)*	1–2	0.71	1.66 (0.17)

[a]*Two items were removed due to low factor loadings;* [b]*One item was removed due to low factor loading;* [c]*This scale was replaced with a single item;* [d]*One item was removed to increase scale reliability.*

Research Misconduct includes:
Fabrication – making up data or results and reporting them
Falsification – manipulating research materials or processes, or changing or omitting data
Questionable research practices – failing to obtain approval, not obtaining consent before conducting research, ignoring outliers, publishing *post hoc* analyses without reporting it, or publishing articles using data that has not been collected legitimately or that has been reported elsewhere.

Procedure

Ethics approval was received from Curtin University Human Research Ethics Committee. Participants were recruited on campus, from a psychology student participant pool and online through social networking sites. The recruiting materials directed potential participants to a Participant Information Sheet hosted on a university website and then linked to the online questionnaire. Consent was assumed upon submitting the questionnaire. Students recruited through the student participant pool were awarded points for participations and other students were provided with the opportunity to enter a draw to win a $50 Amazon.com gift voucher.

Data for 248 cases was downloaded from curtin.qualtrics.com into SPSS (version 21) for data preparation, and cleaning. Duplicate cases and cases with patterned responses or substantial missing data were removed, leaving 205 cases for analysis. A Missing Values Analysis indicated 0.38% missing data across the questionnaire. Little's MCAR test indicated the data was not missing completely at random: χ^2 (1053, $N = 205$) = 1173.68, $p = 0.006$. Expectation Maximization was used to replace missing values. Items were checked for outliers and unusual cases, and scale items were reverse coded where required. Descriptive norms item 3, "In the past year how many students do you think have engaged in research misconduct and have not been caught," was excluded from further analyses due to wide variability in the types of responses yielded, including precise quantitative estimates (76.55%), vague qualitative estimates, such as "a few" (18.53%) and missing data (4.87%).

Confirmatory Factor Analysis was conducted in EQS 6.1to confirm the factor structure of scales in the augmented Theory of Planned Behavior model. Comparative fit indices, with recommended cut-offs from Kline (2011) were used to evaluate the fit of each scale. Based on poor fit statistics and identification of items with low loadings, the attitudes scale was reduced from six-items to five items and the behavior scale was reduced from nine-items to seven items. Goodness of fit statistics could not be computed for the Subjective norms and Descriptive norms scales, and for these measures Principal Axis Factoring supported one-factor solutions. A low Cronbach's alpha of 0.16 and small positive correlations between items indicated the descriptive norms scale was unsuitable for use. Instead, the single item, "Approximately what percentage of students do you think engage in some kind of research misconduct?" was used to represent descriptive norms. Cronbach's alpha was calculated for each of the measures (see

TABLE 2 | Percentage of participants self-reporting engaging in research misconduct.

Behavior	% Engaged in behavior
Claimed to conduct research that was not actually conducted	10.3
Reported research results without obtaining consent from peers	4.9
Claimed to use research materials that were not actually used	17.6
Fabricated information or research data	14.6
Falsified results	12.2
Concealed poor experiment or research data	16.6
Deliberately provided the wrong references	17.1
Deliberately ignored, concealed or distorted unfavorable research results claims	19.5
Provided references at the wrong place of the assignment	37.2

Table 1). The 12-item original social desirability scale yielded a Cronbach's alpha of 0.69. An examination of the questionnaire item-total statistics indicated an improved alpha of 0.71 if the item, "If you say you will do something, do you always keep your promise no matter how inconvenient it might be?" was deleted. This item was deleted, leaving an 11-item scale.

Results

Descriptive Statistics

There was a low level of engagement in the more serious forms of research misconduct. Analysis at the item level (**Table 2**) indicates that approximately one in seven students reported engaging in fabrication and one in eight students in falsification. The proportion of students engaging in questionable research practices varied by type of practice. In total, 39.5% of students admitting to engaging in at least one form of research misconduct (including questionable research practices) at least once.

A summary of descriptive statistics for each measure is presented in **Table 1**. Descriptive norms, intention, justifications, social desirability, and behavior were positively skewed and subjective norms negatively skewed. Analyses were conducted with and without transformations of variables, however, as the results were approximately equivalent the results of the untransformed data are presented for ease of interpretation.

Age, gender, and years of study were not significantly associated with research misconduct behavior and were dropped as control variables. Only social desirability was significantly related to behavior and was retained as the sole control variable for further analyses.

Testing the Augmented Theory of Planned Behavior Model

Prior to commencing analysis, assumptions underlying mediation (Baron and Kenny, 1986) were tested in the correlation matrix (**Table 3**). The criterion variable (behavior), mediators (intention and justification) and predictors (attitude, subjective norms, descriptive norms, perceived behavioral control) were significantly correlated, meeting the requirements for mediation testing. A partial correlation matrix was computed, to control for the effects of social desirability. Path analysis was conducted using LISREL software to enable the simultaneous assessment of all pathways in the model. The testing was conducted in stages. Fit statistics for each stage of testing are presented in **Table 4**.

In the first stage, a partial mediation model was tested. The direct pathways between attitudes and behavior and subjective norms and behavior were non-significant, consistent with the fully mediated relationship in the posited model. However, in contrast to the posited model, the direct pathway between perceived behavioral control and behavior was non-significant, indicating perceived behavioral control is fully mediated by intention and justifications. Also in contrast to the posited model, there was a significant direct pathway between descriptive norms and behavior, indicating a partially, rather than fully mediated relationship.

In the second stage, the model was rerun with the non-significant pathways between attitudes and behavior and subjective norms and behavior removed. All remaining pathways were significant. The predictor variables (attitudes, subjective norms, descriptive norms, and perceived behavioral control) accounted with 23.5% of the variance in intention and 25.6% of the variance in justifications. Intention, justifications and descriptive norms accounted for 38.7% of the variance in behavior. Modification indices indicated a pathway from justification to intent would improve model fit. This pathway is plausible as it is likely that viewing research misconduct behaviors as justifiable would precede the formation of intent to engage in those behaviors.

In the third stage, the pathway from justification to intent was added and the model rerun. With the pathway added, the four other predictors of intent (attitudes, subjective norms, descriptive norms and perceived behavioral control) were no longer significant, suggesting that a simplified model was required, with the relationship between the four predictors and intent fully mediated by justifications.

In the fourth stage, this revised model (**Figure 2**) with the relationships between predictor variables and justification mediated by intent, was tested. This model accounted for 25.6% of the variance in justifications, 50.7% of the variance in intent and 40.8% of the variance in behavior. The Chi Square test was non-significant and fit statistics indicated good model fit to the data. While some fit statistics are superior for the model in the third stage of testing, the final revised model is preferred as it presents a more parsimonious model with good fit statistics and no non-significant pathways.

Discussion

This research examined the Theory of Planned Behavior model, augmented by descriptive norms and justification, in predicting student engagement in research misconduct and questionable research practices. Model testing identified a parsimonious two step mediation model provided good fit to the data. The effect of predictor variables (attitudes, subjective norms,

TABLE 3 | Pearson's correlations between model and control variables.

	1	2	3	4	5	6	7	8	9	10	11	12
(1) Attitudes	1											
(2) SN	−0.42***	1										
(3) DN	0.02	−0.13	1									
(4) PBC	0.09	−0.06	0.17*	1								
(5) Intention	0.33***	−0.34***	0.25***	0.24**	1							
(6) Justification	0.32***	−0.37***	0.24***	0.28***	0.71***	1						
(7) Behavior	0.24**	−0.23**	0.30***	0.21**	0.63***	0.52***	1					
(8) SD	0.00	0.08	0.11	0.08	0.04	0.03	−0.16*	1				
(9) Gender	−0.10	0.04	0.10	−0.19**	−0.15*	−0.06	−0.03	0.10	1			
(10) Age	−0.18**	0.19**	−0.05	0.12	−0.14*	−0.19**	−0.12	0.06	0.02	1		
(11) Yrs of Stdy	−0.05	0.12	0.14*	0.22**	−0.13	−0.08	0.02	−0.08	0.10	0.28***	1	
(12) Und/Post	0.20	0.09	0.05	0.17*	−0.10	0.02	0.12	−0.08	0.03	0.14*	0.37***	1

*SN, subjective norms; DN, descriptive norms; PBC, perceived behavioral conrol; SD, social desirability; Und/Post, undergraduate/postgraduate; Yrs of Stdy, years of study; *p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 4 | Fit indices for models tested.

Model testing	X^2 **sig**	**CFI**	**NNFI**	**SRMSR**	**RMSEA**	**AIC**	**BIC**
Recommended value	***p* > 0.05**	**≥0.9[a]**	**≥0.9[a]**	**<0.1[b]**	**≤0.05[a]**	**lowest**	**lowest**
Stage 1 Partially mediated model	$p < 0.001$	0.79	−3.47	0.10	0.69	1197	1287
Stage 2 Three pathways removed	$p < 0.001$	0.79	−0.08	0.10	0.34	1192	1271
Stage 3 Pathway from justification to intent added	$p = 0.93$	1.00	1.039	0.01	0.0	1095	1178
Stage 4 Revised model	$p = 0.13$	0.99	0.98	0.04	0.05	1097	1163

X^2*, chi square test; CFI, comparative fit index; NNFI, non-normed fit index; SRMSR, standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike information criterion; BIC = Bayesian information criterion;* [a]*Benet-Martínez and Karakitapoglu-Aygün (2003);* [b]*Marsh et al. (2004).*

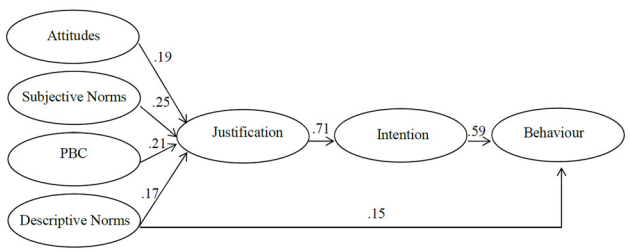

FIGURE 2 | Revised model of theory of planned behavior augmented by descriptive norms and justifications.

descriptive norms, and perceived behavioral control) on behavior was mediated by justifications, with justifications in turn mediated by intention. This revised augmented model accounted for a substantial 40.8% of the variance in student engagement in research misconduct and questionable research practices.

Examination of individual pathways indicates that attitudes, subjective norms, descriptive norms, and perceived behavioral control combined influence intent to engage in research misconduct and questionable research practices through informing the development of justifications (accounting for just over a quarter of the variance in justifications). Justifications accounted for more than half the variance in intent, highlighting the important role of justifications in intent to engage in research misconduct. This is consistent with previous research findings of the important role of justifications/rationalizations/neutralizations in shaping academic dishonesty (e.g., Haines et al., 1986; Labeff et al., 1990; Rettinger and Kramer, 2009; Meng et al., 2014).

As hypothesized, the effect of attitudes and subjective norms on behavior was fully mediated by justification and intention, although the effect of descriptive norms on behavior was only partially mediated. Contrary to the hypothesized partial mediation relationship, the effect of perceived behavioral control on behavior was fully mediated by justification and intention.

These results demonstrate the utility of the augmented Theory of Planned Behavior model in predicting student engagement in research misconduct and questionable research practices. The addition of justifications to the model helps explain the relationship between predictor variables and intent when predicting these dishonest behaviors. The results indicate that viewing research misconduct and questionable research practices positively, believing significant others to also view these positively, perceiving other students to be engaged in these dishonest behaviors and perceiving engaging in these behaviors as easy are associated with justifying engagement in research misconduct and questionable research practices, leading to greater intent and extent of involvement in research misconduct and questionable research practices. However, as this study is cross-sectional it is not possible to establish the causal direction of these findings. It is possible that, as proposed by Stone et al. (2009) in relation to academic misconduct, cognitive dissonance resulting from

engagement in research misconduct and questionable research practices has resulted in individuals trivializing or amending their cognitions in order to reduce dissonance. The addition of descriptive norms increased the predictive ability of the Theory of Planned Behavior model, contributing directly to the prediction of student engagement in research misconduct and questionable research practices and indirectly through justifications. These findings are consistent with previous research findings indicating the importance of observing others' dishonest behavior (Rettinger and Kramer, 2009) and support the utility of adding descriptive norms (Rivis and Sheeran, 2003; Forward, 2009; Stone et al., 2009, 2010; White et al., 2009) and justifications (Stone et al., 2009) to the Theory of Planned Behavior model.

Subjective and descriptive norms were differentially associated with intention, justifications and behavior. Subjective norms were more strongly associated with intention ($r = -0.34$) and justifications ($r = -0.37$) than behavior ($r = -0.23$), while descriptive norms were more strongly associated with behavior ($r = 0.30$) than intention ($r = 0.25$) or justifications ($r = 0.24$). While both types of norms were predictors of intention, only descriptive norms was predictive of behavior once other variables were controlled. These findings support Fishbein and colleagues' recommendation to model both injunctive and descriptive norms within studies (Fishbein, 2000; Fishbein and Yzer, 2003; Ajzen and Fishbein, 2005), and are consistent with meta-analytic results indicating the relationship between descriptive norms and behavior is stronger than the relationship between subjective norms and behavior (Manning, 2009).

In this study ∼40% of students admitting to engaging in at least one form of research misconduct at least once, with one in seven reporting engaging in data fabrication and one in seven engaging in falsifying results. Falling within the lower range of previous estimates of the prevalence of student research misconduct (Franklyn-Stokes and Newstead, 1995; Lawson et al., 1999/2000; Davidson et al., 2001; Brimble and Stevenson-Clarke, 2005; McCabe, 2005), these results confirm that engagement in research misconduct is not restricted to the 'hard sciences,' but is also present to some degree in other disciplines such as psychology.

The consistently reported student engagement in research misconduct and questionable research practices across studies highlights the need to address this type of dishonest behavior in undergraduate and postgraduate programs. The revised augmented Theory of Planned Behavior model increases our understanding of the routes to student engagement in research misconduct and questionable research practices and can be used to identify potential strategies to address these behaviors in universities. Attitudes were a significant predictor of justifications for engaging in research misconduct and questionable research practices. Explicit teaching in research methods courses about resultant harms from these behaviors may help foster a climate where research misconduct is viewed as unacceptable. For example, Boskovic et al. (2013) trialed discussion groups on research misconduct with Ph.D. students. The role of research mentors (Wocial, 1995; Wright et al., 2008; Kornfield, 2012) and supervisors (Mitchell and Carroll, 2008) in educating students about research integrity has also been stressed. However, it has been noted that mentors can exert both positive and negative influence in relation to research misconduct and questionable research practices (Anderson et al., 2007). Fostering a climate that values research integrity may also change subjective and descriptive norms over time.

A further avenue for reducing student engagement in research misconduct and questionable research practices is to directly address the justifications used to reduce ethical dissonance prior to engaging in these behaviors. Removing justifications for dishonest behavior reduces the likelihood of engaging in the behavior (Shalvi et al., 2012). Justifications may be addressed through increasing ethical salience and reducing ambiguity (Shalvi et al., 2015). Ethical salience can be increased through reference to moral codes and standards (Mazar et al., 2008). Further, previous research has indicated that signing a statement of honesty before self-reporting increases ethical salience and reduces dishonest reporting, in comparison to signing after self-reporting. Applying these findings to student research, students could be asked to sign a statement agreeing to engage in ethical research practices as outlined in relevant research ethics codes and guidelines prior to collecting or analyzing data. While completion and signing of ethics applications may serve this function for dissertation students, many lower level student research exercises do not have a requirement to complete and submit an ethics application. As part of the process of removing justifications, any ambiguity surrounding the acceptance of research misconduct and questionable research practices needs to be addressed. In particular, clarity is required on the body of behaviors referred to as 'questionable research practices,' with even the term itself suggesting ambiguity in whether or not these research practices are ethically acceptable. Teaching staff and research supervisors need to provide clear guidance to students on what is, and is not, acceptable research practice, providing applied disciplinary examples.

Perceived behavioral control was also a significant predictor of justifications, indicating that measures could be put in place to make it more difficult to engage in research misconduct and questionable research practices, or at least increase the perception that this type of dishonest behavior is likely to be identified. Procedures have already been developed to detect fabrication of data (Blasius and Thiessen, 2012), and these procedures have now been applied to detecting fabrication in honors dissertations (Allen et al., 2015). In the same way that students are currently required to submit work for plagiarism detection, it is possible in the future that students could be required to submit data-sets for fabrication detection.

Limitations and Future Research

There are a number of limitations of this research that mean caution is required in the interpretation of these results. First, the descriptive norms measure had poor internal reliability, and an individual item providing ratio data was used in its place. This item was predictive of both justifications and behavior, indicating its importance and warranting further development of a descriptive norms measure for use in future research. Second, some variables exhibited non-normality and heteroscedasticity, violating assumptions underlying the analyses. However, analyses using transformed and untransformed data produced similar

results, providing confidence in our findings. Third, self-report measures of past research misconduct and questionable research practices were used as a proxy for future behavior. While this is a common practice in Theory of Planned Behavior research (Armitage and Conner, 2001), future research separating the time of measurement of intention and behavior is recommended. This is particularly important when justifications are included in the model, as it has been argued that justifications may be made based on previous engagement in misconduct (Stone et al., 2009). Fourth, the reliance on self-report methods for all variables introduces the risk of common method variance/bias. However, recent *post hoc* research examining the effect of common method variance on Theory of Planned Behavior studies has indicated that common method variance is not a concern within this domain (Schaller et al., 2015). The reliance on self-report measures is also likely to have resulted in under-reporting of behavior (see John et al., 2012 for comparison of prevalence rates of questionable research practices with and without incentives for honesty in responding). Despite this, self-reports of engaging in research misconduct and questionable research practices provide a useful indicator of these behaviors. Previous research has demonstrated associations between self-reports of dishonest behavior and actual engagement in dishonest behaviors (Halevy et al., 2014), increasing our confidence in their use as proxies for actual behaviors. Finally, the majority of students in this study were psychology students from one university, limiting the generalizability of these findings to other academic settings. We recommend future research is based on larger samples across disciplines and universities, enabling a stronger test of the hypotheses. The actual and perceived seriousness and consequences of research misconduct and questionable research practices may vary according to student level and type of research project (e.g., assignment versus dissertation) and larger samples will enable an assessment of both the prevalence of these behaviors and the validity of the model by year group.

Conclusion

In this research the Theory of Planned Behavior model, augmented by descriptive norms and justification, was used to predict student engagement in research misconduct and questionable research practices. The results support a two-step mediation model, where the effect of attitudes, subjective norms, descriptive norms and perceived behavioral control on behavior is mediated first by justifications, and then intention. The model has good utility, able to account for 40% of the variance of student engagement in research misconduct and questionable research practices.

References

Ajzen, I. (1985). "From intentions to actions: a theory of planned behavior," in *Action Control: From Cognition to Behaviour*, eds J. Kuhl and J. Beckmann (Berlin: Springer), 11–39.

Ajzen, I. (1991). The theory of planned behaviour. *Organ. Behav. Hum. Decis. Process.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T

Ajzen, I., and Fishbein, M. (2005). "The influence of attitudes on behaviour," in *The Handbook of Attitudes*, eds D. Albarracín, B. T. Johnson, and M. P. Zanna (Mahwah, NJ: Lawrence Erlbaum), 173–221.

Allen, P. A., Laurenco, A., and Roberts, L. D. (2015). Detecting students' research data fabrication: A method and illustration. *Ethics. Behav.* doi: 10.1080/10508422.2015.1019070

Alleyne, P., and Phillips, K. (2011). Exploring academic dishonesty among university students in Barbados: an extension to the theory of planned behaviour. *J. Acad. Ethics* 9, 323–338. doi: 10.1007/s10805-011-9144-1

Anderson, M. S., Horn, A. S., Risbey, K. R., Ronning, E. A., De Vries, R., and Martinson, B. C. (2007). What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. *Acad. Med.* 82, 853–860. doi: 10.1097/ACM.0b013e31812f764c

Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E., and Kamata, T. (2013). "Research integrity and misconduct in the academic profession," in *Higher Education: Handbook of Theory and Research*, Vol. 28, ed. M. Paulsen (Berlin: Springer), 217–261. doi: 10.1007/978-94-007-5836-0_5

Ariely, D. (2012). *The (Honest) Truth About Dishonesty: How We Lie to Everyone-Especially Ourselves*. London: Harper Collins. doi: 10.1177/0972262912483993

Armitage, C. J., and Conner, M. (2001). Efficacy of the theory of planned behaviour: a meta-analytic review. *Br. J. Soc. Psychol.* 40, 471–499. doi: 10.1348/014466601164939

Arvidson, C. J. (2004). *The Anatomy of Academic Sishonesty: Cognitive Sevelopment, Self-Concept, Neutralization Techniques, and Attitude Toward Cheating.* (*Doctoral Dissertation*). Available at: http://www.researchgate.net/publication/34492252

Bandura, A. (1978). Social learning theory of aggression. *J. Commun.* 28, 12–29. doi: 10.1111/j.1460-2466.1978.tb01621.x

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *J. Exp. Psychol. Gen.* 141, 757–773. doi: 10.1037/a0027588

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Beck, L., and Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behaviour. *J. Res. Pers.* 25, 285–301. doi: 10.1016/0092-6566(91)90021-H

Bedian, A. G., Taylor, S. G., and Miller, A. N. (2010). Management science on the credibility bubble: cardinal sins and various misdemeanors. *Acad. Manage. Educ.* 9, 715–725. doi: 10.5465/AMLE.2010.56659889

Bennett, R. (2005). "Factors associated with student plagiarism in a post-1992 university," in *Assess. Eval. High. Educ.* 30, 137–162. doi: 10.1080/0260293042000264244

Benet-Martínez, V., and Karakitapoglu-Aygün, Z. (2003). The interplay of cultural syndromes and personality in predicting life satisfaction: comparing Asian Americans and European Americans. *J. Cross Cult. Psychol.* 34, 38–61. doi: 10.1177/0022022102239154

Blasius, J., and Thiessen, V. (2012). *Assessing the Quality of Research Data*. London: Sage. doi: 10.1080/0260293042000264244

Boskovic, M., Djokovic, J., Grubor, I., Guzvic, V., Jakovljevic, B., Juresivic, M., et al. (2013). PhD students' awareness of research misconduct. *J. Empir. Res. Hum. Res. Ethics* 8, 163–164. doi: 10.1525/jer.2013.8.2.163

Brimble, M., and Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *Aust. Educ. Res.* 32, 19–44. doi: 10.3200/JRLP.138.2.101-114

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015

Claxton, L. D. (2005). Scientific authorship. Part 1. A window into scientific fraud? *Mutat. Res.* 589, 17–30. doi: 10.1016/j.mrrev.2004.07.003

Davidson, E. W., Cate, H. E., Lewis, C. M., and Hunter, M. (2001). "Data manipulation in the undergraduate laboratory: what are we teaching?" in *Proceedings of the First ORI Research Conference of Research Integrity*, Bethesda, MD: Office of Research Integrity, 209–213.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4:1–11. doi: 10.1371/journal.pone.0005738

Fang, F. C., Steen, R. G., and Casadevell, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17028–17033. doi: 10.1073/pnas.1212247109

Festinger, L. A. (1957). *A Theory of Cognitive Dissonance*. Evanston: Stanford University Press. doi: 10.1073/pnas.1212247109

Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care* 12, 273–278. doi: 10.1080/09540120050042918

Fishbein, M., and Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Commun. Theory* 13, 164–183. doi: 10.1111/j.1468-2885.2003.tb00287.x

Forward, S. E. (2009). The theory of planned behaviour: the role of descriptive norms and past behaviour in the prediction of drivers' intention to violate. *Transp. Res. Part F Traffic Psychol. Behav.* 12, 198–207. doi: 10.1016/j.trf.2008.12.002

Francis, L. J., Brown, L. B., and Philipchalk, R. (1992). The development of an abbreviated form of the revised Eysenck personality questionnaire (EPQR-A): it's use among students in England, Canada, the U.S.A. and Australia. *Pers. Individ. Dif.* 13, 443–449. doi: 10.1016/0191-8869(92)90073-X

Franklyn-Stokes, A., and Newstead, S. E. (1995). Undergraduate cheating: who does what and why? *Stud. High. Educ.* 20, 159–172. doi: 10.1080/03075079512331381673

Gilbert, F. J., and Denison, A. R. (2003). Research misconduct. *Clin. Radiol.* 58, 499–504. doi: 10.1016/S0009-9260(03)00176-4

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gottfredson, M. R., and Hirshci, T. (1990). *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Graves, S. M. (2008). Student cheating habits: a predictor of workplace deviance. *J. Divers. Manage.* 3, 15–22.

Grieneisen, M. L., and Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLoS ONE* 7:e44118. doi: 10.1371/journal.pone.0044118

Haines, V. J., Diekhoff, G. M., LaBeff, E. E., and Clark, R. E. (1986). College cheating: immaturity, lack of commitment, and the neutralizing attitude. *Res. High. Educ.* 25, 342–354. doi: 10.1007/BF00992130

Halevy, R., Shalvi, S., and Verschuere, B. (2014). Being honest about dishonesty: correlating self-reports and actual lying. *Hum. Commun. Res.* 40, 54–72. doi: 10.1111/hcre.12019

Harding, T. S., Mayhew, M. J., Finelli, C. J., and Carpenter, D. D. (2007). The theory of planned behaviour as a model of academic dishonesty in engineering and humanities undergraduates. *Ethics. Behav.* 17, 255–279. doi: 10.1080/10508420701519239

Hofman, B., Myr, A. I., and Holm, S. (2013). Scientific dishonesty – a nationwide survey of doctoral students in Norway. *BMC Med. Ethics* 14:3. doi: 10.1186/1472-6939-14-3

Jann, B., Jerke, J., and Krumpal, I. (2012). Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. *Public Opin. Q.* 76, 32–49. doi: 10.1093/poq/nfr036

John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23, 524–532. doi: 10.1177/0956797611430953

Kallgren, C. A., Reno, R. R., and Cialdini, R. B. (2000). A focus theory of normative conduct: when norms do and do not affect behavior. *Pers. Soc. Psychol. B.* 26, 1002–1012. doi: 10.1177/01461672002610009

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*, 3rd Edn. New York, NY: The Guildford Press. doi: 10.1177/0956797611430953

Kornfield, D. S. (2012). Research misconduct: the search for a remedy. *Acad. Med.* 87, 877–882.

Kumar, M. N. (2008). A review of the types of scientific misconduct in biomedica l research. *J. Acad. Ethics* 6, 211–228. doi: 10.1007/s10805-008-9068-6

Labeff, E. E., Clark, R. E., Haines, V. J., and Diekhoff, G. M. (1990). Situational ethics and college student cheating. *Sociol. Inq.* 60, 190–198.

Lawson, A. E., Lewis, C. M. Jr., and Birk, J. P. (1999/2000). Why do students "cook" data? A case study on the tenacity of misconceptions. *J. Coll. Sci. Teach.* 29, 191–198.

Manning, M. (2009). The effects of subjective norms on behaviour in the theory of planned behaviour: a meta-analysis. *Brit. J. Soc. Psychol.* 48, 649–705. doi: 10.1348/014466608X393136

Marsden, H., Carroll, M., and Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Aust. J. Psychol.* 57, 1–10. doi: 10.1080/00049530412331283426

Marsh, H. W., Hau, K.-T. and Wen, Z. (2004). In search of golden rules: comments on hypothesis-testing approaches to setting cuitoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct. Eq. Modeling* 11, 320–341. doi:10.1207/s15328007sem1103_2

Martinson, B. C., Anderson, M. S., and de Vries, R. (2005). Scientists behaving badly. *Nature* 435, 737–738. doi: 10.1038/435737a

Mavrinac, M., Brumini, G., Bilic-Zulle, L., and Petrovecki, M. (2010). Construction and validation of attitudes toward plagiarism questionnaire. *Croat. Med. J.* 51, 195–201. doi: 10.3325/cmj.2010.51.195

Mayhew, M. J., Hubbard, S. M., Finelli, C. J., Harding, T. S., and Carpenter, D. D. (2009). Using structural equation modeling to validate the theory of planned behavior as a model for predicting student cheating. *Rev. High. Educ.* 32, 441–468. doi: 10.1353/rhe.0.0080

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

McCabe, D. L. (2005). Cheating among college and university students: a North American perspective. *Int. J. Educ. Integr.* 1, 1–11.

McCabe, D. L., Feghali, T., and Abdallah, H. (2008). Academic dishonesty in the Middle East: individual and contextual factors. *Res. High. Educ.* 49, 451–467. doi: 10.1007/s11162-008-9092-9

McCabe, D. L., and Trevino, L. K. (1993). Academic dishonesty: honor codes and other contextual influences. *J. Higher Educ.* 64, 522–538.

Meng, C. L., Othman, J., D'Silva, J. L., and Omar, Z. (2014). Influence of neutralization attitude in academic dishonesty among undergraduates. *Int. Educ. Stud.* 7, 66–73. doi: 10.5539/ies.v7n6p66

Mitchell, T., and Carroll, J. (2008). Academic and research misconduct in the PhD: issues for students and supervisors. *Nurse Educ. Today* 28, 218–226. doi: 10.1016/j.nedt.2007.04.003

National Health and Medical Research Council and Australian Research Council. (2007). *Australian Code for the Responsible Conduct of Research*. Information Report, Commonwealth of Australia, Canberra, ACT. doi: 10.1016/j.nedt.2007.04.003

Nonis, S., and Swift, C. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: a multi-campus investigation. *J. Educ. Bus.* 77, 69–77. doi: 10.1080/08832320109599052

Ogilvie, J., and Stewart, A. (2010). The integration of rational choice and self-efficacy theories: a situational analysis of student misconduct. *Aust. N. Z. J. Criminol.* 43, 130–155. doi: 10.1375/acri.43.1.130

Park, C. (2003). In other (people's) words: plagiarism by university students – literature and lessons. *Assess. Eval. High. Educ.* 28, 471–488. doi: 10.1080/0260293032000120352

Pickard, J. (2006). Staff and student attitudes to plagiarism at university college Northampton. *Assess. Eval. High. Educ* 31, 215–232. doi: 10.1080/02602930500262528

Pimple, K. D. (2002). Six domains of research ethics: a heuristic framework for the responsible conduct of research. *Sci. Eng. Ethics* 8, 191–205. doi: 10.1007/s11948-002-0018-1

Reno, R. R., Cialdini, R. B., and Kallgren, C. A. (1993). The transsituational influence of social norms. *J. Pers. Soc. Psychol.* 64, 104–112. doi: 10.1037/0022-3514.64.1.104

Rettinger, D. A., and Kramer, Y. (2009). Situational and personal causes of student cheating. *Res. High. Educ.* 50, 293–313. doi: 10.1007/s11162-008-9116-5

Rise, J., Sheeran, P., and Hukkelberg, S. (2010). The role of self-identity in the theory of planned behavior: a meta-analysis. *J. Appl. Soc. Psychol.* 40, 1085–1105. doi: 10.1111/j.1559-1816.2010.00611.x

Rivis, A., and Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: a meta-analysis. *Curr. Psychol.* 22, 218–233. doi: 10.1007/s12144-003-1018-2

Rose, L. L. (2008). *Scientific Misconduct: Perceptions, Beliefs, Working Environments, and Reporting Practices in the Clinical Research Associate Population.* (*Doctoral Dissertation*). Available at: http://books.google.com.au/

Schaller, T. K., Patil, A., and Malhotra, N. K. (2015). Alternative techniques for assessing common method variance an analysis of the theory of planned behavior research. *Organ. Res. Methods* 18, 177–206. doi: 10.1177/1094428114554398

Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychol. Sci.* 23, 1264–1270. doi: 10.1177/0956797612443835

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications: doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24, 125–130.

Soper, D. S. (2013). *A-priori Sample Size Calculator for Hierarchical Multiple Regression*. Available at: http://www.danielsoper.com/statcalc3/calc.aspx?id=16

Steen, R. G. (2010). Retractions in the scientific literature: do authors deliberately commit research fraud? *J. Med. Ethics* 15, 1–5. doi: 10.1136/jme.2010.038125

Steen, R. G. (2011). Retractions in the scientific literature: is the incidence of research fraud increasing? *J. Med. Ethics* 37, 249–253. doi: 10.1136/jme.2010.040923

Steneck, N. H. (2006). Fostering integrity in research: definitions, current knowledge and future directions. *Sci. Eng. Ethics* 12, 53–74. doi: 10.1007/PL00022268

Stone, T. H., Jawahar, I. M., and Kisamore, J. L. (2009). Using the theory of planned behavior and cheating justifications to predict academic misconduct. *Career Dev. Int.* 14, 221–241. doi: 10.11.08/13620430910966415

Stone, T. H., Kisamore, J. L., and Jawahar, I. M. (2010). Predicting academic misconduct intentions and behavior using the theory of planned behavior and personality. *Basic Appl. Soc. Psych.* 32, 35–45. doi: 10.1080/01973530903539895

Sykes, G. M., and Matza, D. (1957). Techniques of neutralization: a theory of delinquency. *Am. Sociol. Rev.* 22, 664–670. doi: 10.1108/13620430910966415

Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics*, 5th Edn. Boston, MA: Pearson/Allyn and Bacon. doi: 10.1080/01973530903539895

US Department of Health and Human Services Office of Research Integrity. (2000). *Federal Research Misconduct Policy* (*Information Report 2000*). Washington, DC: Office of Science and Technology Policy. doi: 10.2307/2089195

White, K. M., Smith, J. R., Terry, D. J., Greenslade, J. H., and McKimmie, B. M. (2009). Social influence in the theory of planned behaviour: the role of descriptive, injunctive and in-group norms. *Brit. J. Soc. Psychol.* 48, 135–158.

Wocial, L. D. (1995). The role of mentors in promoting integrity and preventing scientific misconduct in nursing research. *J. Prof. Nurs.* 11, 276–280.

Wright, D. E., Titus, S. L., and Cornelison, J. B. (2008). Mentoring and research misconuct: an analysis of research mentoring in closed ORI cases. *Sci. Eng. Ethics* 14, 323–336. doi: 10.1007/s11948-008-9074-5

Yang, S. C. (2012a). Attitudes and behaviors related to academic dishonesty: a survey of Taiwanese graduate students. *Ethics Behav.* 22, 218–237. doi: 10.1080/10508422.2012.672904

Yang, S. C. (2012b). Ethical academic judgments and behaviors: applying a multidimensional ethics scale to measure the ethical academic behavior of graduate students. *Ethics Behav.* 22, 281–296. doi: 10.1080/10508422.2012.672907

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 21 July 2015
doi: 10.3389/fpsyg.2015.00877

Social-cognitive barriers to ethical authorship

Jordan R. Schoenherr[1, 2]*

[1] Department of Psychology, Carleton University, Ottawa, ON, Canada, [2] Department of Innovation in Medical Education, University of Ottawa, Ottawa, ON, Canada

Keywords: research misconduct, research integrity, inappropriate authorship, source credibility, applied ethics

Introduction

The apparent increase in research misconduct in the scientific literature has caused considerable alarm in both the biomedical (Benos et al., 2005; Smith, 2006) and psychological research communities (Stroebe et al., 2012). An understanding of research misconduct must be informed by the recognition that the norms of science might be quite general (e.g., Merton, 1942; Bronowski, 1965), ambiguous (Cournand and Meyer, 1976), or even contradictory (e.g., Mitroff, 1974; Ziman, 2000), leading to possible disagreements in terms of what constitutes misconduct within a research community (Fields and Price, 1993; Berk et al., 2000; Al-Marzouki et al., 2005). Considerable insight can be gained from research on behavioral ethics (e.g., Bazerman and Tenbrunsel, 2011; Ariely, 2012; Greene, 2013). Using inappropriate authorship practices as an illustrative example, I consider the role of social-cognitive mechanisms in research misconduct while also suggesting preventative measures.

OPEN ACCESS

Edited by:
Shahar Ayal,
Interdisciplinary Center Herzliya, Israel

Reviewed by:
Guy Hochman,
Duke University, USA
Andrea Pittarello,
Ben-Gurion University of the Negev, Israel

***Correspondence:**
Jordan R. Schoenherr,
jordan.schoenherr@carleton.ca

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *05 March 2015*
Accepted: *13 June 2015*
Published: *21 July 2015*

Citation:
Schoenherr JR (2015) Social-cognitive barriers to ethical authorship. Front. Psychol. 6:877. doi: 10.3389/fpsyg.2015.00877

Prevalence of Research Misconduct

Widespread interest in dishonesty in research began comparatively recently in the history of the sciences (e.g., Broad and Wade, 1982; Steneck, 1999) although there was an early recognition that misconduct was a feature of scientific research (Babbage, 1830). Though a definitive set of forms of misconduct has yet to be identified, fabrication, falsification, and plagiarism (FFP) are generally cited as clear violations of scientific norms. In a review of studies of FFP, Steneck (2006) estimated that its occurrence rate fell within a range of 1.0 and 0.001% (for recent support, see Fanelli, 2009). He further suggested that research practices reflect a normal distribution, with FFP representing outlying behaviors. More ambiguous behaviors, or questionable research practices (QRP), have a much higher rate of occurrence, with Steneck suggesting that they constitute 10–50% of all research practices. QRPs represent an interesting form of misconduct in that they apparently reflect a feature of normal science (De Vries et al., 2006) thereby suggesting that they might reflect the social-cognitive processes underlying the dishonest behaviors of people more generally (e.g., Bazerman and Tenbrunsel, 2011; Ariely, 2012).

Inappropriate authorship practices are a prevalent form of QRP. For instance, they can represent a failure to recognize an original contribution to research (*ghost authorship*) or a misattribution of the research to those who have not contributed (*gift authorship*). The prevalence of inappropriate authorship practices is reflected in studies conducted by Flanagin et al. (1998) and Wislar et al. (2011) wherein they observed a decrease in the prevalence of ghost authorship from 11.5 to 7.9% between 1996 and 2008. In contrast, the number of articles affected only by gift authorship remained relative constant with a non-significant decrease from 19.3 to 17.6% during the same period (for similar findings, see Mowatt et al., 2002; Mirzazadeh et al., 2011; cf. Stretton, 2014). Accounting for the stability and change of inappropriate authorship practices represents an important task for applied ethics as the assignment of

credit can lead to stratification within the scientific community (e.g., Cole and Cole, 1973).

The Social Cognition of Credit and Credibility

Early commentators attributed research misconduct to a range of factors including publication pressure, competition, and psychopathy (Chubin, 1985; cf. Braxton and Bayer, 1994). However, the prevalence of QRP suggests that more general social-cognitive mechanisms can account for research misconduct. Analyses of cases of misconduct have suggested a number of contributing factors (for a review, see Davis et al., 2007). Here I will consider how inappropriate authorship practices can be understood in terms of influence of social conventions and conformity, the reciprocity norms of exchange systems, as well as role schemata and status.

Social Conventions and Conformity Bias

The social conventions and ethical norms of science are evidenced in its cultural, structural, and organizational systems (Davis, 2003). Empirical support for the role of social conventions in judgements of ethical conduct comes from a number of sources. Kohlberg (1976) outlines a model with three stages of moral reasoning. A preconventional stage of moral reasoning defined by self-interest is contrasted against a subsequent stage of conventional moral reasoning wherein social norms of the group or society are used to judge behavior. While an additional post-conventional stage relies on the use of ethical principles, Kohlberg found that few individuals achieve this stage of reasoning (cf. Rest et al., 1999). Even when morals can be clearly identified, conventions play an important role in social interactions (Turiel, 2002) with conformity biases maintaining cultural norms (e.g., Whiten et al., 2005; Efferson et al., 2008). Experimental evidence also suggests that dishonest behaviors increase when in-group members are observed to engage in these behaviors (Gino et al., 2009).

Studies of academic misconduct have also demonstrated the influence of conventions and conformity, in terms of peer influence on cheating. In their study, McCabe and Treviño (1997) found that peer behavior and fraternity/sorority membership were positively related to the occurrence of misconduct, whereas perceived peer disapproval was negatively related to the occurrence of misconduct (see also, McCabe et al., 2001). Social conventions additionally offer an explanation for the difficulty in implementing successful ethics training programs, with disciplinary and departmental values being associated with researcher behavior (e.g., Anderson et al., 1994) and regression from post-conventional reasoning to conventional reasoning (Rennie and Rudland, 2002; Hren et al., 2011).

Social Organization and Reciprocal Exchange

The nature and prevalence of dishonesty can also be understood in terms of the norms of social exchange systems (e.g., Fiske, 1991). Fiske (1991) considers four kinds of exchange systems that differ in terms of the commensurability of the objects in the reciprocal exchange relationship (*equality matching*; *communal sharing*; *market pricing*; and *authority ranking*). These systems will in turn determine what is seen as honest and dishonest behavior. For instance, a researchers' contributions to a research project (e.g., theory, data collection, statistics) might be deemed unique and incommensurable, making judgments of proportion of credit arbitrary (*communal sharing*) or exceedingly difficult (*equality matching*). Researchers might instead assume that contributions can be differentiated and are quantifiable in terms of an absolute value that can be used to assign a proportion of authorship credit and responsibility (*market pricing*). Rightly or wrongly, this exchange norm appears to underscore the belief that the order of authorship reflects the proportion of contribution a researcher has made to a study (e.g., ICJME, 2005/2008). Finally, researchers might assume that authority should be the primary determinant of the assignment of credit (*authority ranking*), something that I will return to the next section.

Scientific research has been defined as an exchange system by a number of authors. Hagstrom (1982) suggested that a research article can be viewed as analogous to a gift whereas Street et al. (2010) have noted that "journal articles are valuable intellectual property," (p. 1458). These observations as well as others suggest that reciprocity can exert considerable influence on our judgements (Gouldner, 1960; Fiske, 1991). In terms of authorship, credit might be given due to the need for reciprocity by junior researchers receiving funding or advice from senior researchers. Authorship deals, or "mutual support authorships," wherein researchers include names of authors so as to have their name included on a project, also explicitly reflect an overt reciprocity strategy (Claxton, 2005; Louis et al., 2008). In addition to overt pressure, "lab chiefs" might be assigned undue credit as a result of researchers receiving career advice and financial support thereby enabling the research process while not directly contributing to intellectual content of a specific publication (Broad and Wade, 1982; Claxton, 2005; Street et al., 2010). Similarly, the provision of sponsorship might be perceived as sufficient grounds for receiving authorship (Louis et al., 2008). Both of these behaviors might be best understood in terms of the *halo effect* (Thorndike, 1920; Nisbett and Wilson, 1977) wherein participants overgeneralize from one attribute to the individual as a whole (see also, Harvey et al., 2010).

Source Credibility, Status, and Role Schemata

Due to the need to allocate limited attention, researchers must identify a subset of individuals that appear to provide credible information (Thorngate et al., 2011). Source credibility exerts considerable influence in the formation and change of attitudes (e.g., Petty et al., 1997). Thus, the contributions of researchers who are deemed to have greater credibility *a priori* might not be judged as critically as those with less credibility. Supporting this, studies that manipulate power (e.g., Guinote, 2013) have demonstrated that those in comparatively powerless position have reduced attention and short-term memory resources due to a need to respond to those in positions of power. In comparison, those in powerful positions are more likely to engage in confirmation bias in the pursuit of their goals. Collaborations between senior and junior researchers will likely be influenced

by these situational factors (e.g., Sullivan and Ogloff, 1998) making it harder for junior members to assess the contributions of senior authors. Gift authorship can also be understood as an instance of a desire to confer credibility onto a research project. Peters and Ceci (1982) demonstrated this influence in a quasi-experiment wherein journal articles previously published by prestigious authors were resubmitted with fictitious non-prestigious names. When submitted with non-prestigious names, the majority of referees rejected these previously accepted articles.

The effects of source credibility can also be understood in terms of status assigned to social roles (e.g., Merton, 1968; Azoulay et al., 2014). Role schemata contain information pertaining to behaviors and obligations associated with a given role in a particular social context, thereby influencing the behavior and judgments of self and others. Historically, Shapin (1989) has noted that despite significant intellectual contributions to the design and conduct of experiments, technicians were not deemed to warrant authorship. As noted above, lab chiefs also appear to be awarded undue credit (Broad and Wade, 1982) and this might be attributed to perceived differences in credibility. If students and other personnel associated with a research project are believed to have a "supporting" role, their contributions might not be attributed to them. Rather, they might need to be legitimated by credible others in order for them to be accepted within a research community. More generally, authority ranking exchange systems assume that those in positions of authority are deemed to warrant more resources (Fiske, 1991). This would manifest itself as being awarded a disproportionate amount of credit. However, role schemata can also benefit those perceived to be in a subordinate position. As Zuckerman (1968) observed, Nobel laureates often appear to have awarded greater authorship credit to less prestigious collaborators. Moreover, those with higher status have also been found to express more favorable attitudes toward preserving the ethical norms of their discipline (e.g., Braxton and Bayer, 1994).

Conclusions

If inappropriate authorship practices can be accounted for by general social-cognitive processes, then an ameliorative program at least appears possible in principle. In opposition to these efforts, ethics training programs developed in an applied context have not always been successful (e.g., Brown and Kalichman, 1998; Fisher et al., 2009). Such failures likely stem from an ethical "fudge factor," a failure to attend to ethical norms on a moment-to-moment basis, and the observation of dishonest behavior of peers (e.g., Bazerman and Tenbrunsel, 2011; Ariely, 2012; Greene, 2013). Indeed, rather than engaging in an explicit reasoning process (Kohlberg, 1976; Rest et al., 1999) our responses to ethical dilemmas often appear to be automatic (Haidt, 2007) and are susceptible to loss framing and time pressure (e.g., Kern and Chugh, 2009). Together with self-deception and justifications (Tenbrunsel and Messick, 2004; Shalvi et al., 2011), ethical facets of authorship decisions might become less salient. Reciprocity norms, along with the "publish or perish" framing of contemporary academic publishing, would certainly support these behaviors. These enablers must be acknowledged and addressed if we hope to reduce ghost and gift authorship.

Having recognized the influence of social context and automaticity, three general proposal appear to offer promise to reduce the prevalence of unethical behaviors. First, we must ensure that researchers are aware of the ethical standard and norms of authorship within their research community and that co-authors discuss expectations and roles throughout the research process. Standards such as those provided by the ICJME (2005/2008) are useful points of reference for the assignment of authorship/contributorship. Second, by continually priming these norms with ongoing discussions at departmental and disciplinary levels, we are likely to obtain similar reductions in dishonest behavior as those observed in laboratory studies (Mazar et al., 2008). Finally, to disincentivize dishonest behavior stemming from a "publish or perish" academic culture, we must consider adopting criterion for hiring, promotion, and funding decisions based on the quality of a restricted number of publications rather than the total number of publications produced by an individual.

Funding

This research was supported by funding from the Ottawa Health Research Institute.

References

Al-Marzouki, S., Roberts, I., Marshall, T., and Evans, S. (2005). The effect of scientific misconduct on the results of clinical trials: a delphi study. *Contemp. Clin. Trials* 26, 331–337. doi: 10.1016/j.cct.2005.01.011

Anderson, M. S., Louis, K. S., and Earle, J. (1994). Disciplinary and departmental effects on observations of faculty and graduate student misconduct. *J. High. Educ.* 65, 331–350.

Ariely, D. (2012). *The Honest Truth about Dishonesty: How We Lie to Everyone, Especially Ourselves*. New York, NY: HarperCollins.

Azoulay, P., Stuart, T., and Wang, Y. (2014). Matthew: effect or fable? *Manage. Sci.* 60, 92–109. doi: 10.1287/mnsc.2013.1755

Babbage, C., (1830/2004). *Reflections on the Decline of Science in England, and on Some of Its Causes*. London: Kessinger Publishing Company.

Bazerman, M. H., and Tenbrunsel, A. E. (2011). *Blind Spots: Why We Fail to Do What's Right and What to Do About It*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400837991

Benos, D. J., Fabres, J., Farmer, J., Gutierrez, J. P., Hennessy, K., Kosek, D., et al. (2005). Ethics and scientific publication. *Adv. Physiol. Educ.* 29, 59–74. doi: 10.1152/advan.00056.2004

Berk, R. A., Korenman, S. G., and Wenger, N. S. (2000). Measuring consensus about scientific research norms. *Sci. Eng. Ethics* 6, 315–340. doi: 10.1007/s11948-000-0035-x

Braxton, J. M., and Bayer, A. E. (1994). Perceptions of research misconduct and an analysis of their correlates. *J. High. Educ.* 65, 351–372. doi: 10.2307/2943972

Broad, W. J., and Wade, N. (1982). *Betrayers of the Truth*. New York, NY: Simon and Schuster.

Bronowski, J. (1965). *Science and Human Values, Rev Edn*. New York, NY: Harper Torchbooks.

Brown, S., and Kalichman, M. W. (1998). Effects of training in the responsible conduct of research. *Sci. Eng. Ethics* 4, 487–498. doi: 10.1007/s11948-998-0041-y

Chubin, D. E. (1985). Misconduct in research: an issue of science policy and practice. *Minerva* 23, 175–202. doi: 10.1007/BF01099941

Claxton, L. D. (2005). Scientific authorship. Part 2. History, recurring issues, practices, and guidelines. *Mutat. Res.* 589, 31–45. doi: 10.1016/j.mrrev.2004.07.002

Cole, J. R., and Cole, S. (1973). *Social Stratification in Science*. Chicago, IL: University of Chicago Press.

Cournand, A., and Meyer, M. (1976). The Scientist's Code. *Minerva* 14, 79–96. doi: 10.1007/BF01096215

Davis, M. S. (2003). The role of culture in research misconduct. *Account. Res.* 10, 189–201. doi: 10.1080/714906092

Davis, M. S., Riske-Morris, M., and Diaz, S. R. (2007). Causal factors implicated in research misconduct: evidence from ORI case files. *Sci. Eng. Ethics* 13, 395–414. doi: 10.1007/s11948-007-9045-2

De Vries, R., Anderson, M. S., and Martinson, B. C. (2006). Normal misbehavior: scientists talk about the ethics of research. *J. Empir. Res. Hum. Res. Ethics* 1, 43–50. doi: 10.1525/jer.2006.1.1.43

Efferson, C., Lalive, R., Richerson, P. J., McElreath, R., and Lubell, M. (2008). Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evol. Hum. Behav.* 29, 56–64. doi: 10.1016/j.evolhumbehav.2007.08.003

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4:e5738. doi: 10.1371/journal.pone.0005738

Fields, K. L., and Price, A. R. (1993). Problems in research integrity arising from misconceptions about the ownership of research. *Acad. Med.* 68, S60–S64. doi: 10.1097/00001888-199309000-00036

Fisher, C. B., Fried, A. L., and Feldman, L. G. (2009). Graduate socialization in the responsible conduct of research: a national survey on the research ethics training experiences of psychology doctoral students. *Ethics Behav.* 19, 496–518. doi: 10.1080/10508420903275283

Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York, NY: Free Press.

Flanagin, A. Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., et al. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA* 280, 222–224. doi: 10.1001/jama.280.3.222

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychol. Sci.* 20, 393–398. doi: 10.1111/j.1467-9280.2009.02306.x

Gouldner, A. W. (1960). The norm of reciprocity: a preliminary statement. *Am. Sociol. Rev.* 25, 161–178. doi: 10.2307/2092623

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York, NY: The Penguin Press.

Guinote, A. (2013). "Social power and cognition," in *The Oxford Handbook of Social Cognition,* ed D. E. Carlston (New York, NY: Oxford University Press), 575–587. doi: 10.1093/oxfordhb/9780199730018.013.0028

Hagstrom, W. O. (1982). "Gift giving as an organizing principle in science," in *Science in Context: Readings in the Sociology of Science*, eds B. Barnes and D. O. Edge (Cambridge: MIT Press), 21–34.

Haidt, J. (2007). The new synthesis in moral psychology. *Science* 316, 998–1002. doi: 10.1126/science.1137651

Harvey, A., Kirk, U., Denfield, G. H., and Read, P. (2010). Monetary favors and their influence on neural responses and revealed preference. *J. Neurosci.* 30, 9597–9602. doi: 10.1523/JNEUROSCI.1086-10.2010

Hren, D., Marušić, M., and Marušić., A. (2011). Regression of moral reasoning during medical education: combined design study to evaluate the effect of clinical study years. *PLoS ONE*, 6:e17406. doi: 10.1371/journal.pone.0017406

International Committee of Medical Journal Editors (ICJME). (2005/2008). *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication*. Available online at: www.icmje.org on November, 2008.

Kern, M., and Chugh, D. (2009). Bounded ethicality: the perils of loss framing. *Psychol. Sci.* 20, 378–384. doi: 10.1111/j.1467-9280.2009.02296.x

Kohlberg, L. (1976). "Moral stages and moralization: the cognitive-development approach," in *Moral Development and Behavior: Theory, Research and Social Issues*, ed T. Lickona (New York, NY: Holt, Rinehart and Winston), 31–53.

Louis, K. S., Holdsworth, J. M., Anderson, M. S., and Campbell, E. G. (2008). Everyday ethics in research: translating authorship guidelines into practice in the bench sciences. *J. Higher Educ.* 79, 88–112. doi: 10.1353/jhe.2008.0002

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

McCabe, D. L., and Treviño, L. K. (1997). Individual and contextual influences on academic dishonesty: a multicampus investigation. *Res. High. Educ.* 38, 379–396.

McCabe, D. L., Treviño, L. K., and Butterfield, K. D. (2001). Cheating in academic institutions: a decade of research. *Ethics Behav.* 11, 219. doi: 10.1207/S15327019EB1103_2

Merton, R. K. (1942). "The normative structure of science," in *The Sociology of Science: Theoretical and Empirical Investigations*, ed R. K. Merton (Chicago, IL: University of Chicago Press), 267–278.

Merton, R. K. (1968). The Matthew effect in science. *Science* 159, 56–63. doi: 10.1126/science.159.3810.56

Mirzazadeh, A., Navadeh, S., Rokni, M. B., and Farhangniya, M. (2011). The prevalence of honorary and ghost authorships in Iranian biomedical journals and its associated factors. *Iran J. Public Health* 40, 15–21. Available online at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3481730/

Mitroff, I. I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: a case study of the ambivalence of scientists. *Am. Sociol. Rev.* 39, 579–595. doi: 10.2307/2094423

Mowatt, G., Shirran, L., Grimshaw, J. J. M., Rennie, D., Flanagin, A., Yank, V. et al. (2002). Prevalence of honorary and ghost authorship in Cochrane reviews. *J. Am. Med. Assoc.* 287, 2769–2771. doi: 10.1001/jama.287.21.2769

Nisbett, R. E., and Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *J. Pers. Soc. Psychol.* 35, 250–256. doi: 10.1037/0022-3514.35.4.250

Peters, D., and Ceci, S. (1982). Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav. Brain Sci.* 5, 187–195. doi: 10.1017/S0140525X00011183

Petty, R. E., Wegener, D. T., and Fabrigar, L. R. (1997). Attitudes and attitude change. *Annu. Rev. Psychol.* 48, 609–647. doi: 10.1146/annurev.psych.48.1.609

Rennie, S. C., and Rudland, J. R. (2002). Differences in medical students' attitudes to academic misconduct and reported behaviour across the years: a questionnaire study. *J. Med. Ethics* 29, 97–102. doi: 10.1136/jme.29.2.97

Rest, J., Narvaez, D., Bebeau, M., and Thoma, S. (1999). A Neo-Kohlbergian approach: the DIT and schema theory. *Educ. Psychol. Rev.* 11, 291–324. doi: 10.1023/A:1022053215271

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Shapin, S. (1989). The invisible technician. *Am. Sci.* 77, 554–563.

Smith, R. (2006). Research misconduct: the poisoning of the well. *J. R. Soc. Med.* 99, 232–237. doi: 10.1258/jrsm.99.5.232

Steneck, N. (1999). Confronting misconduct in science in the 1980s and 1990s: What has and has not been accomplished? *Sci. Eng. Ethics* 5, 161–175. doi: 10.1007/s11948-999-0005-x

Steneck, N. (2006). Fostering integrity in research: definitions, current knowledge, and future directions. *Sci. Eng. Ethics* 12, 53–74. doi: 10.1007/s11948-006-0006-y

Street, J. M., Rogers, W. A., Israel, M., and Braunack-Mayer, A. J. (2010). Credit where credit is due? Regulation, research integrity and the attribution of authorship in the health sciences. *Soc. Sci. Med.* 70, 1458–1465. doi: 10.1016/j.socscimed.2010.01.013

Stretton, S. (2014). Systematic review on the primary and secondary reporting of the prevalence of ghostwriting in the medical literature. *BMJ Open* 4:e004777. doi: 10.1136/bmjopen-2013-004777

Stroebe, W., Postmes, T., and Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspect. Psychol. Sci.* 7, 670–688. doi: 10.1177/1745691612460687

Sullivan, L. E., and Ogloff, J. R. P. (1998). Appropriate supervisor-graduate student relationships. *Ethics Behav.* 8, 229–248. doi: 10.1207/s15327019eb0803_4

Tenbrunsel, A. E., and Messick, D. M. (2004). Ethical fading: the role of self-deception in unethical behavior. *Soc. Justice Res.,* 17, 223–236. doi: 10.1023/B:SORE.0000027411.35832.53

Thorndike, E. L. (1920). A constant error in psychological ratings. *J. Appl. Psychol.* 4, 25–29.

Thorngate, W., Liu, J., and Chowbhury, W. (2011). The competition for attention and the evolution of science. *J. Artif. Soc. Soc. Simul.* 14, 1–6. Available online at: http://jasss.soc.surrey.ac.uk/14/4/17.html

Turiel, E. (2002). *The Culture of Morality: Social Development, Context, and Conflict*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511613500

Whiten, A., Horner, V., and de Waal, F. B. M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature* 437, 737–740. doi: 10.1038/nature 04047

Wislar, J. S., Flanagin, A., Fontanarosa, P. B., and DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross-sectional survey. *Br. Med. J.* 343:d6128. doi: 10.1136/bmj. d6128

Ziman, J. (2000). *Real Science: What it is and What it Means*. Cambridge: Cambridge University Press.

Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers: a study of social symbolism and its ambiguity. *Am. J. Sociol.* 74, 276–291. doi: 10.1086/224641

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on