

**Judging Those We Cheat:
How Committing Unethical Behavior affects
Subsequent Judgments of its Victims**

Thesis for the degree of "Doctor of Philosophy"

by

Nurit Hod

Submitted to the Senate of the Hebrew University
of Jerusalem
September/2021

This work was carried out under the supervision of
Professor Eyal Pe'er

Acknowledgments

I would like to express my deepest gratitude and appreciation to my advisor, Prof. Eyal Peer. Your guidance, continuous support, extensive knowledge, and constructive criticism have inspired me to become a better researcher and made my doctoral journey productive and memorable. It has been an honor and a privilege working with you.

I would also like to thank my dissertation committee: Prof. Shai Danziger and Prof. Shahar Ayal, for their helpful guidance, insightful comments, and willingness to help at all times. Special thanks to Prof. Shai. Your knowledge and input were invaluable and much appreciated. I am also grateful to my peers, former and present Ph.D. students. I would especially like to thank Hila Schupak. Your friendship and support have contributed to my sanity through this long and lonely process.

Finally, I would like to thank my family: My parents for their support throughout this endeavor; My mother-in-law, for her encouragement and for always being there for me; My wonderful children, Eitam and Lotem, who endured many hours of absence. Thank you for your patience and understanding in the past few years. Last, but certainly not least, special thanks to my life partner and true love, Avi. You were there every step of the way, in every win and every loss, in the long days and even longer nights. Thank you for walking hand-in-hand with me on this journey and making sure that I am protected and loved. I feel both grateful and blessed.

Abstract

Unethical behavior is a widespread phenomenon, with high financial costs to organizations and society at large. Acts such as shoplifting, using company resources for personal use, tax evasion, and insurance fraud are some examples of acts that ordinary people commit regularly and that, eventually, result in significant economic impact. Research on unethical behavior has focused mainly on the antecedents of such acts and how they can be mitigated. However, research has typically overlooked the possible impact of unethical behavior on subsequent judgments of the target of the transgression. For example, if self-employed underreport their income for taxes, would it influence their judgment of tax-related authorities? If asked, would they provide a more positive or negative evaluation of the government? Current theories suggest conflicting predictions for such questions. On the one hand, people have been shown to experience high degrees of guilt following unethical acts, which could lead to reparative actions and prosocial behavior, suggesting that perpetrators of unethical behavior would judge the target of their transgression more favorably. On the other hand, when people find it difficult to deny or dismiss their wrongdoing, they may distance themselves from the act by criticizing others to restore their moral self-image. This suggests that perpetrators of unethical behavior might judge the target of their transgression more negatively.

In this research, I contrast these opposing predictions to better understand the important downstream consequence of dishonesty—namely, people’s judgment and reviews of those they have cheated. Also, I explore whether and how subsequent judgment is affected by the ability to justify unethical behavior. Lastly, I also examine the impact of unethical behavior on written review's nature and sentiments and the possible influence they have on other people's behaviors and willingness to engage with the victim. The phenomenon was examined in different settings, with both real-life cases and online performance. Findings suggest that when people behave unethically, they judge the victim more harshly, presumably to distance themselves, and thus justify their transgression. The results were consistent so that employees judged their employers less favorably after recalling unethical behaviors in the workplace, and self-employed judged tax-related government authorities more negatively after recalling their false tax reports. Similarly, in marketing settings, participants who tested a new app judged it more negatively the more they cheated. Also, demonstrating an important boundary condition, I found that unethical behavior does not influence judgments when cheating was blatantly obvious and hard to justify. This finding supports a distancing-based process as the underlying

mechanism driving the effect. Finally, I found that when people behave unethically, they tend to use more negative sentiments in their written reviews and reduce readers' willingness to engage with the target of the transgression. However, reviews generated by participants who were given the opportunity to cheat, but curbed their dishonesty despite the temptation, increased readers' willingness to use the product compared to reviews generated by participants who could not cheat.

The research has a significant theoretical contribution since it highlights an unexplored aspect of unethical behavior. The research considers the impact unethical behavior has on subsequent judgments and evaluations rather than focus on antecedents and ways to mitigate it. That is, the research considers unethical behavior as the cause rather than the effect. Additionally, the research contributes to the growing field of word of mouth by revealing a novel factor that can systematically sway ratings and written reviews. Since unethical behavior is prevalent, many online reviews likely originate from unethical behavior. Findings suggest that these reviews may be biased due to the tendency of people who behave immorally to harshly judge the victim of their transgression.

In addition, the research has important implications for organizations, policymakers, and marketers because of the lingering negative effects unethical behavior has. Since organizations are subject to public opinion, negative judgments of unethical people may influence their image in the public's eyes, affect business results, and consequently their ability to recruit good employees, provide quality services, or keep the stakeholders' interests. Findings also emphasize the importance of policies that aim to deter people from behaving unethically, since firms that use lenient policies may expose themselves to more negative online reviews. Lastly, the research suggests implications concerning consumer protection and fair-trade policies. Since unethical behavior has significant consequences in terms of word of mouth, people should be informed about possible biases that may influence their decision-making while considering a product or interacting with organizations. Lastly, policymakers should also consider policies that aim to protect organizations' interests due to the negative consequences of unethical behaviors.

A Letter of Contribution

This dissertation is written in the form of a compilation of articles (unpublished; The articles will be submitted for publication in academic journals). All the chapters were written by the student Nurit Hod under the supervision and guidance of Prof. Eyal Pe'er.

Nurit Hod (student)

A handwritten signature in blue ink, consisting of stylized capital letters 'N' and 'H' followed by a horizontal line.

Prof. Eyal Pe'er (supervisor)

A handwritten signature in black ink, written in a cursive style as 'Eyal Pe'er'.

Contents

Introduction.....	9
Overview.....	9
Antecedents of Unethical Behavior	11
Self-Justification's Role in Unethical Behavior	11
Consequences of Unethical Behavior.....	13
Impact on Subsequent Judgment.....	14
Research Overview	16
Chapter 1 The Impact of Unethical Behavior on Subsequent Judgments of Organizations....	19
Abstract.....	19
Introduction.....	20
Unethical Behavior in Organizational Settings	20
Regulations and Interventions – Shift in Policies.....	21
Subsequent Judgment and Possible Impact on Policies	23
Study 1.1 - Pretest	25
Study 1.2 - Impact of perceived severity and prevalence of questionable behaviors on job satisfaction	27
Study 1.3 - Impact of engaging in questionable behaviors on job satisfaction	31
Study 1.4 - Tax evasion and judgment of tax authorities	35
General Discussion	39
References.....	43
Chapter 2 Consumer's Unethical Behavior, Self-Justification, and Subsequent Judgments ...	48
Abstract.....	48
Introduction.....	49
Unethical Consumption	49
Subsequent Judgment in Marketing Realm.....	51
Distancing as a Self-Justification Mechanism	53
Study 2.1 – How unethical consumers judge a product.....	54
Study 2.2 – The moderating role of ease of justification.....	59
Study 2.3 – The moderating role of moral credentials.....	62

Study 2.4 - The moderating role of actively benefiting from cheating: pretest.....	67
Study 2.5 – The moderating role of actively benefiting from cheating	69
General Discussion	74
References.....	79
Chapter 3 The Impact of Reviews by Unethical Individuals on Other People's Decisions	84
Abstract.....	84
Introduction.....	85
Online Reviews	85
Written Reviews and Sentiments.....	87
The Impact of Unethical Behavior on written Reviews	88
Study 3.1 – How unethical consumers review a product.....	91
Study 3.2 - The effect of reviews from ethical/unethical consumers on the willingness to use the product.....	100
General Discussion	105
References.....	109
Discussion and Conclusions	116
References..	123

Introduction

"People hasten to judge in order not to be judged themselves. What do you expect? The idea that comes most naturally to man, as if from his very nature, is the idea of his innocence."

- Albert Camus, *The Fall*

Overview

Unethical behavior is a common, everyday phenomenon. Shoplifting, tax evasion, and over-claiming insurance are just some examples of acts that ordinary people commit regularly (e.g., Ayal & Gino, 2011). Although many ethical violations are minor, enabling people to self-justify them (Shalvi, Eldar, & Bereby-Meyer, 2012), their economic impact is significant, accounting for billions of dollars annually. For example, recent data suggest that employees' unethical behavior is estimated at 5% of firm revenues and translates into \$4.5 trillion at the global level (ACFE, 2020); Tax non-compliance is estimated at \$406 billion in lost funding annually in the United States (IRS, 2016). In addition, insurance fraud accounts for 10% of all claim payments, or \$32 billion annually (Insurance Information Institute, 2016), and illegal software downloads totaled \$46.3 billion globally in 2018 alone (Business Software Alliance, 2018). Notably, unethical behavior in organizations and the marketplace result in a 10 to 15 percent increase in the price of consumer goods, costing American families billions of dollars a year (Feldman, van Rooij, & Rorie, 2019). Dishonesty also threatens a company's reputation and its ability to offer quality services (Singh & Twalo, 2015).

Given the significant costs of unethical behavior to society, research has mostly focused on the scope and causes of unethical behavior (Ayal, Hochman, & Ariely, 2016; Gerlach, Teodorescu, & Hertwig, 2019). Traditional economic theories have emphasized the external cost-benefit analysis people presumably make (Becker, 1968), while traditional psychological approaches focused on people's ethical judgment and attitudes as driving misconducts (Hunt & Vitell, 2006; Muncy & Vitell, 1992). In recent years, the focus has shifted to behavioral decision theories and approaches that aim to explain the underlying causes of unethical behavior, specifically through identifying the mechanisms that allow people to engage in unethical behavior without threatening their moral self-image (Mazar, Amir, & Ariely, 2008; Schurr, Ritov, Kareev, & Avrahami, 2012; Shalvi, Gino, Barkan, & Ayal, 2015).

Despite the breadth of research on what drives unethical behavior, much less research has examined how unethical acts affect actors' subsequent behaviors and judgment. The limited work which has been done has examined the effects of unethical behavior on post-cheating emotional reactions (e.g., Ruedy, Moore, Gino, & Schweitzer, 2013), guilt and guilt-relief (Cohen, Wolf, Panter, & Insko, 2011; Ruedy et al. 2013; Xu, Bègue, & Bushman, 2012), subsequent prosocial behavior (Gneezy, Imas, & Madarász, 2012), and how money that is obtained illegally is spent (Levav & McGraw, 2009). Critically, however, these studies all focus on the effects of unethical behavior on people's subsequent behavior *outside* the context of their transgression and do not consider how it can impact people's behavior and judgment toward those they have wronged.

Unethical acts, such as cheating on tax reports, or shoplifting (e.g., by purposefully not scanning an item in a self-checkout cashier), lead to significant financial losses but may also result in additional unexplored behavioral consequences that could affect the victim. For example, consider a shopper who uses a self-checkout cashier at the supermarket and shoplifts by purposefully not scanning an item. Will this unethical act influence the shopper's judgment of the supermarket? When asked to evaluate the supermarket, will having "cheated" impact the shopper's evaluations? Existing theories have not considered these implications and actually provide conflicting predictions of unethical behavior's subsequent judgment. On the one hand, unethical acts may cause guilt that triggers reparative actions and prosocial behavior (Gneezy et al., 2012; Xu et al., 2012), suggesting that the shoplifter would judge and review the supermarket more favorably to reduce feelings of guilt. On the other hand, research shows that people often distance themselves from their own misconduct, preferring to criticize others so they can view themselves as "ultra-moral" (Barkan, Ayal, Gino, & Ariely, 2012; Shalvi et al., 2015). This account suggests that the aforementioned shoplifter would judge and review the supermarket more negatively to distance themselves from the object of the misconduct.

Obviously, these different predictions cannot hold simultaneously, and the current research aims to solve this apparent theoretical contradiction by examining how cheaters judge those they have cheated. However, before addressing any possible behavioral consequences, it is critical to review the antecedents of unethical behavior in order to highlight the financial benefits and psychological costs cheaters face, which can eventually impact their subsequent behavior.

Antecedents of Unethical Behavior

The basic theoretical model used in nearly all research on unethical behavior begins with the economics-of-crime model of Becker (1968), which emphasized the external cost-benefit analysis people presumably make. The traditional economic approach to ethical violations assumes that people first evaluate the expected benefits (rewards or gains) and costs (potential punishment, informal sanctions, combined with the probability of detection) of the unethical act and then determine their course of action. Moreover, the rational agent is assumed to compare the expected costs and benefits of unethical activity with those of legitimate activities and rationally allocate time and resources to one of the alternatives. However, empirical studies found that predictions of the rational choice theory are sometimes unsupported (Van Winden & Ash, 2012) and emphasized the need to focus on people's ethical judgment in determining beliefs, attitudes, and actual misconducts (Hunt & Vitell, 2006; Muncy & Vitell, 1992).

According to theories in the consumer ethics field, ethical judgment is determined by two stages of moral evaluations of the possible behaviors. In the first stage of the evaluation, the person weights behavior according to applied norms or personal values, whereas in the second stage of the evaluation, the person weights the consequences of each behavior (Hunt & Vitell, 2006). However, as consistently found in different domains, there is often a gap between people's attitudes and their actual behavior (De Pelsmacker, Driesen, & Rayp, 2005; Sheeran, 2002; Sniehotta, Scholz, & Schwarzer, 2005). That is, upon examining the consequences and norms relative to a particular behavior, a person might perceive it to be the most ethical but still adopt a different alternative.

To address this gap, new theories have emerged in recent years, with more attention paid to behavioral aspects of unethical behavior and the justifications people use in their decision-making process. Self-justification, whether used before or after engaging in unethical behavior, serves to excuse unethical behavior as morally acceptable and thus allow people to preserve moral self-image and avoid psychological costs of it.

Self-Justification's Role in Unethical Behavior

Abundant evidence suggests that people do not always maximize their gains when cheating and, despite the temptation, tend to curb their own dishonesty (Ayal, Gino, Barkan, & Ariely, 2015). Even when they know their unethical behavior will never be revealed, people show a certain level of aversion toward cheating, limiting it to an extent far below the maximum possible (Fischbacher & Föllmi-Heusi, 2013; Hilbig & Hessler, 2013; Mazar et al., 2008;

Pruckner & Sausgruber, 2006; Shalvi, Dana, Handgraaf, & De Dreu, 2011). This tendency to profit from unethical behavior while limiting it to preserve an honest self-image remains consistent even when people are presented with the explicit risk of getting caught (Gamliel & Peer, 2013).

The two opposing desires people seem to hold - to profit from illegal acts and see themselves as moral, can lead to an “ethical dissonance”. Ethical dissonance is a feeling of uncomfortable tension due to the inconsistency between one's actual dishonest behavior and the personal values associated with one's moral self-concept (Ayal & Gino, 2011; Barkan et al., 2012). When people face an anticipated or experienced ethical dissonance, they tend to justify their misbehavior and excuse it as morally acceptable or not entirely unethical (Barkan et al., 2012; Shalvi et al., 2015).

Justifications for unethical behavior reinforce a sense of consistency between the behavior in question and desired moral standards and reduce moral dissonance (Kunda, 1990; Shalvi et al., 2015) by providing reasons for the questionable act, emphasizing extenuating circumstances, or softening the moral standard of the person (Barkan et al., 2012). People can use pre-violation justifications to excuse their questionable behavior and lessen the anticipated threat to their moral self by redefining their questionable behaviors as excusable. However, if the unethical behavior has already been committed, people may use post-violation justifications to compensate for their misbehavior and reduce the experience of ethical dissonance (Shalvi et al., 2015).

One of the common situations in which people tend to use justifications before engaging in unethical behavior is while facing ambiguity. When norms and rules are imprecise and open to multiple interpretations, people can frame the unethical behavior as less immoral (Schweitzer & Hsee, 2002). Blurred criteria for judging what is right or wrong may allow people the opportunity to reinterpret their unethical behavior and whitewash it (Ayal & Gino, 2011; Barkan et al., 2012). However, even when the rules are clear, people may find a way to frame the situation as morally ambiguous. For example, a study showed that people tend to over-report the number of the first (paid) roll of a die when given the possibility to roll the die three times instead of one time (Fischbacher & Föllmi-Heusi, 2013). While reporting a high number after one roll requires lying by inventing a number, adding two irrelevant rolls allows people to inflate the number they received in the first roll if it did appear on a later roll. That is, people can frame the situation as more ambiguous and, therefore, be more flexible in inventing facts (Shalvi et al., 2015).

Another justification that can be used before acting immorally is moral licensing, which serves as a "moral balance scale". When people feel sufficiently moral (having collected a surplus of moral credentials), they may feel licensed to refrain from good behavior and act immorally (Ayal & Gino, 2011; Effron, Cameron, & Moni, 2009; Miller & Effron, 2010; Monin & Miller, 2001; Shalvi et al., 2015). For example, in making a consumption decision, people were more likely to take reckless purchase decisions after being asked to imagine having volunteered to do community service (Khan & Dhar, 2007). Similarly, after disclosing a conflict of interest to their clients, advisors sometimes felt licensed to provide biased advice (Cain, Loewenstein, & Moore, 2005).

While moral licensing allows people to restrict moral behavior and return to a more comfortable level, in cases where moral self-image is threatened after engaging in unethical acts, people can reduce their ethical dissonance by engaging in moral cleansing (Ayal & Gino, 2011). Moral cleansing serves as a way of restoring moral self-image (Ayal & Gino, 2011; Tetlock, Kristel, Elson, Green, & Lerner, 2000) by allowing people to clear themselves from the immoral act and to turn a new page in their moral ledger (Sachdeva, Iliev, & Medin, 2009). Moral cleansing can take a physical or symbolic form - moral statements, prosocial intentions, physical washing, inflicting pain to one's self, or confessions (Barkan et al., 2012; Peer, Acquisti, & Shalvi, 2014; Zhong & Liljenquist, 2006). While moral licensing provides justifications *before* the unethical behavior, moral cleansing helps deal with a transgression *after* the act.

Considering self-justification mechanisms, whether before or after people engage in unethical activity, enables a better understanding of people's behavior in various contexts and how they may profit from it while avoiding psychological costs. However, most studies have focused on identifying predictors of unethical behavior rather than the outcomes and consequences of these activities (Ruedy et al., 2013). Thus, it is still unclear how self-justification can be used to predict and explain the consequences of unethical behavior.

Consequences of Unethical Behavior

Studies suggest that behaving immorally can negatively influence the perceptions of self-worth and positive self-image (Sachdeva et al., 2009). Therefore, people behave unethically in a limited manner: enough to increase their profit but less than the extent threatening their positive self-image as honest and decent people (Ayal & Gino, 2011; Hilbig & Hessler, 2013; Mazar et al., 2008). Since people desire to perceive themselves as honest and deserving, they strongly

believe in their own morality (Aquino & Reed, 2002), even if doing so requires a certain degree of self-deception or pretense (Barkan et al., 2012). Because of the desire to be moral and to be seen by others as such, engaging in dishonest acts usually result in guilt, a negative and self-conscious emotion (Gneezy et al., 2012; Shu & Gino, 2012; Tracy & Robins, 2004, 2007). The self-evaluation process evokes guilt and functions to promote reparative actions such as confessions, apologies, and attempts to undo the harm (Cohen et al., 2011; Tangney, 2003; Tracy & Robins, 2007).

Additionally, in order to repent for their sins and relieve the feeling of guilt, people may also engage in prosocial behaviors (Sachdeva et al., 2009; Xu et al., 2012). Prosocial behaviors include increased willingness to comply with direct requests for help (Shu & Gino, 2012) or donation to a charity (Gneezy et al., 2012). Evidence shows that immoral choices need to be recent, not only in time but also in the extent of the choice's moral outcome, since willingness to engage in prosocial behavior following unethical behavior tends to decay over time (Gneezy et al., 2012).

Still, there are cases in which consumers cannot dismiss, deny or confess their misconducts and therefore may experience unresolved tension and guilt. To relieve the unethical dissonance, people tend to use distancing responses in which they judge others' behavior more harshly and present themselves as more virtuous and ultra-honest (Barkan et al., 2012; Shalvi et al., 2015). However, unlike moral cleansing and licensing behaviors that are oriented inward, distancing aims for audience recognition and self-presentation rather than to judgment of the self (Barkan et al., 2012).

Although existing models of (un)ethical decision-making have assumed that unethical behavior triggers negative affect, it seems that in some cases, unethical behaviors may also trigger a positive affect, which is known as a "cheater's high". In a series of experiments, participants who cheated (with no obvious victim) experienced positive affect compared to those who did not cheat. Factors such as financial incentives and rationalization were ruled out with the notion that the cheater's high reflects the thrill of getting away with cheating (Ruedy et al., 2013).

Impact on Subsequent Judgment

If unethical behavior can elicit a wide array of emotions and behaviors, unethical behavior should also impact people's attitudes and behavior in subsequent situations. Existing theories of unethical behavior can help generate predictions for how cheating might affect subsequent

judgments of the victims of the unethical behavior. Because people desire to be moral and seen by others as such, dishonest acts usually result in guilt (Gneezy et al., 2012; Shu & Gino, 2012). Feeling guilty may encourage reparative actions and prosocial behaviors, such as empathy and care-giving, to “balance the scale” (Cohen et al., 2011; Tracy & Robins, 2007). Thus, people may act positively towards those they have cheated. As noted in the opening example, the shopper who feels guilty for having shoplifted may judge the supermarket more positively and provide better recommendations than had he or she not shoplifted.

However, some researchers note that compensation behavior may occur only if the person fails to resolve the ethical dissonance by justifying the act (Barkan et al., 2012; Shalvi et al., 2015). In other words, unethical individuals may *ex-ante* restrict themselves from engaging in behaviors that are too unjustifiable and therefore will not regard their actions as unethical. As such, cheating will not influence their subsequent judgments. Indeed, self-justification can reduce prosocial behavior's proclivity (Hofmann, Wisneski, Brandt, & Skitka, 2014). Thus, if cheating is justifiable, less guilt arises, and there is less need to compensate the victim in order to relieve the ethical dissonance.

Still, people sometimes find it difficult to dismiss, deny, or otherwise justify their misconduct and, therefore, still experience ethical dissonance. To alleviate that dissonance, people may try to distance themselves from the act by judging the behavior of others more harshly or presenting themselves as more virtuous or ultra-honest (Barkan et al., 2012; Shalvi et al., 2015). In other words, consistent with their moral history and past decisions, people might act more immorally after an initial immoral behavior (Schwabe, Dose, & Walsh, 2018), and therefore continue to judge their victim more negatively to preserve a sense of moral self-consistency (Jussim, Yen, & Aiello, 1995). The tendency to judge the victim more negatively corresponds with theories that imply that people can selectively disengage moral self-regulation (for instance, by attributing the blame to the victim) in order to make unethical acts more socially or morally acceptable (Bandura, 1999).

In conclusion, a considerable amount of research has addressed the issue of unethical behavior in numerous contexts – finance, marketing, taxes, workplaces – while trying to understand why people behave immorally. Most of the recent literature explores the way people maintain their moral self-image by using justifications (before or after engaging in unethical behavior) and accordingly excuse their behavior as less immoral. It appears that although unethical behavior is expected to have a significant impact on subsequent judgments and behaviors, current theories diverge in their prediction of the direction and possibly the

magnitude of these effects. In this dissertation, I aim to contrast these opposing predictions to better understand how people judge the victims of their own unethical behavior.

Research Overview

The current research has three main objectives: first, to examine the important downstream consequence of dishonesty—namely, people’s judgment and reviews of those they have cheated. Subsequent judgment of the victims will be examined in different settings towards different entities including organizations, government authorities, and marketers, with both real-life cases and online performance. The second objective is to explore the moderating role of justification and understand whether subsequent judgment is affected by the ability to justify unethical behavior. In addition, I examine additional boundary conditions such as moral credentials and the degree to which people actively cheated (vs. passive cheating). The third objective is to explore the impact of unethical behavior on written review's nature and sentiments and the possible influence they have on other people's behaviors and willingness to engage with the victim. In addition, I discuss the implications of better understanding the impact of unethical behavior on subsequent judgment while developing effective policies by managers and policymakers.

The following three chapters lay the foundations for a better understanding of the phenomenon and the implications in various contexts:

The first chapter: "The Impact of Unethical Behavior on Subsequent Judgments of Organizations." examines how people judge those they have cheated. More specifically, I explore the way employees and businesspersons who recalled their unethical behaviors toward their workplace or tax authorities judge those institutions. In four studies, I find that people tend to judge the victims of their transgressions more harshly, presumably to distance themselves and thus justify their transgression. These findings contribute to the body of knowledge by confronting the opposing predictions of current theories. In addition, it has practical implications for managers and policymakers in their effort to maintain an organization's image in the minds of their employees and the public.

The second chapter: "Consumer's Unethical Behavior, Self-Justification, and Subsequent Judgments." has two main goals. The first goal is to extend the findings from organizational settings to the marketing realms. The results indicate that the negative consequence of unethical acts is likely to be quite pervasive because it happens not only in long-term relationships, such as with an employer or the government but also in one-time

interactions. One-time interactions often characterize consumer behavior in the marketing realm, such as using a certain product or service. The second goal is to explore the moderation role of the ability to justify the unethical behavior. Demonstrating an important boundary condition, I find that the unethical act does not influence judgments when cheating is relatively obvious and hard to justify.

The third chapter: "How Unethical Behavior Impact Online Written Reviews and Sentiments?" explores the impact of unethical behavior on written review's nature and sentiments and the possible influence they have on other people's behaviors and willingness to engage with the victim. Since people tend to use social media to review products, services, and companies, it is important to understand if unethical behavior impacts the reviews' different aspects. Text analysis reveals that people who cheat tend to use more negative sentiments in the reviews, but only when cheating is not obvious (i.e., not hard to justify). Further examination reveals that reviews written after behaving unethically reduced other people's willingness to engage with the victim. These findings reinforce the importance of word-of-mouth in shaping others' judgments and behaviors and the notion that unethical behavior not only influences the subsequent judgment of the cheater but may affect the perception and image of organizations and government in the public eyes.

Exploring the impact of unethical behavior on subsequent judgments is innovative since it sheds light on unexplored aspects of unethical behavior by shifting the focus from antecedents of unethical behavior and ways to mitigate it to the impact unethical behavior probably has on subsequent judgment. That is, the research considers unethical behavior as the cause rather than the effect. The study also adds to the body of knowledge of online reviews, which up until now have not considered unethical behavior as a possible influencing factor (e.g., Babić, Sotgiu, De Valck, & Bijmolt, 2016; Cheung & Lee, 2012). Since 60% of consumers are estimated to have engaged in unethical behavior at some point in their lives (Krasnovsky & Lane, 1998), it is thus clear that some online reviews may originate from cases of unethical behavior, and the difference between these and ordinary reviews could prove to be significant.

The research may have far-reaching implications for organizations, policymakers, and marketers in terms of business strategies and effective policies. Because cheating is common and people's judgments drive their choices and word of mouth, it may influence other people's decisions and behaviors. Hence, unethical behavior may affect an organization's perception and impact brand equity and business results. As a result, firms and policymakers may decide to

change business or marketing strategies to approach the consequences of unethical behavior, such as shifting resources from expensive preventive measures to more creative marketing efforts along the customer journey. In addition, policies concerning consumer protection and fair trade may need to consider ways to inform the public about possible bias due to unethical behavior in daily life, both in the marketing realm, workplaces, and organizations such as governmental authorities.

In conclusion, this research uncovers an unexplored consequence of unethical behavior, that is, how people judge those they have cheated. Behaving unethically may encourage people to distance themselves and harshly judge organizations and brands, providing lower ratings and potentially causing further harm to those they have already cheated. This research provides important implications for managers and policymakers seeking to maintain the organization's image in the minds of the general public and customers and to protect fair trade principles in general.

Chapter 1

The Impact of Unethical Behavior on Subsequent Judgments of Organizations

Abstract

Although much of individuals' unethical behavior is directed towards organizations, the impact of individuals' dishonesty on their judgment of the organization is unknown. Previous research has neglected to explore the potentially important downstream consequence of cheating. Existing theories offer contradicting predictions: on the one hand, cheaters may feel guilty for having cheated, so they judge the organization more favorably; on the other hand, to protect their self-image, cheaters may blame the organization for their misbehavior, consequently judging it more negatively. In four experiments, I find that cheaters judge organizations more negatively the more they cheat and when cheating behavior is made salient. These findings have important implications for organizations and policymakers because cheating is common, and individuals' judgments drive their own future choices and word of mouth, thus influencing other people's behavior.

Introduction

People have often been found to behave unethically towards organizations: employees can defraud their workplace, citizens can cheat their government, and consumers can poach sellers and vendors. While the media highlight extreme cases of fraud and scams, many ethical violations are minor and committed by ordinary people who value morality but behave unethically when faced with the opportunity to cheat (Gino, 2015). Although these minor ethical violations enable people to self-justify them (Shalvi, Eldar, & Bereby-Meyer, 2012), their aggregate economic impact is enormous, accounting for billions of dollars annually (Business Software Alliance, 2018; National Retail Federation, 2020; Taylor, 2016), and play a significant part in the bankruptcies of organizations (Feldman, van Rooij, & Rorie, 2019).

Given the significant costs of unethical behavior to society, much research has explored the antecedents and potential remedies for unethical behavior (Ayal, Hochman, & Ariely, 2016; Gerlach, Teodorescu, & Hertwig, 2019). However, studies have failed to consider an important downstream consequence of dishonesty—namely, people’s judgment and evaluations of those they have cheated. Because unethical behavior is common (e.g., Ayal & Gino, 2011) and individuals’ judgments drive their own future choices and word of mouth, thus influencing others’ behavior, understanding whether unethical behavior has systematic downstream effects is important.

Unethical Behavior in Organizational Settings

Unethical behavior is a common, everyday phenomenon, especially in organizational settings. Employees inflating business expense reports, committing payroll fraud, overstating performance or contributions to teamwork, stealing inventory or intellectual property, and abusing company assets for personal use are acts that ordinary people regularly commit (Ayal & Gino, 2011; Gill, Prowse & Vlassopoulos, 2013). Employees' unethical behavior in the workplace has reached epidemic proportions throughout the world (Weber, Kurke, & Pentico, 2003), and according to recent data, one out of every 50 employees was caught stealing from their employer in 2019 (Hayes international annual retail theft survey, 2019). The estimated loss due to unethical behaviors is estimated at 5% of firm revenues and translates into \$4.5 trillion at the global level (ACFE, 2020). In addition to the direct impact on business revenue, unethical behavior in the workplace can threaten a company's reputation and its ability to offer quality services to customers and stakeholders (Singh & Twalo, 2015).

Unethical behavior in an organizational setting can also be found among individuals and entrepreneurs who cheat on their tax reports to reduce their legally due tax obligations (Alm & Torgler, 2011). Such actions can include underreporting income, claiming unearned deductions, or failing to file appropriate tax returns. Tax non-compliance is a major concern to government and public policy. It is estimated at \$406 billion in lost funding annually in the United States (IRS, 2016) and £34.1 billion in the UK (HMRC, 2019). Tax evasion impacts governmental revenues and can also lead to significant unfairness in society (Bott, Cappelen, Sørensen, & Tungodden, 2019).

Interestingly, even when people behave unethically, they are usually not perceived as criminals, either legally or socially, and succeed in maintaining their moral self-image as ethical people (Feldman et al., 2019). This tendency has the danger of normalizing individuals' dishonesty and violation of rules in general. The prevalence of dishonest behaviors and the threat to the legitimacy of the organizations' rule systems lead to considerable pressure on policymakers and organizations to effectively monitor and detect unethical behavior (Gill et al., 2013; Singh & Twalo, 2015). Importantly, in recent years there is a shift in policy and interventions. While in the past, measures relied on the notion that people are rational and try to maximize their gains, in recent years, measures are based on behavioral aspects and justifications people use in their decision-making process (Feldman, 2018; Zhang, Gino, & Bazerman, 2014).

Regulations and Interventions – Shift in Policies

Sanctions and penalties are frequently used to discourage people from engaging in unethical behaviors. By forming sanctions, policymakers and managers rely on the idea that a sanction changes the behavior's desirability and therefore deters people from dishonest acts (Mulder, 2018). This view is based on classical approaches that assume people are rational agents who aim to maximize the expected utility of their gains, weighing the benefits of successful cheating against the risky consequences of detection and punishment (Dell'Anno, 2009). However, evidence suggests that classical measures to limiting unethical behavior in organizations, such as formal codes of conduct or high fines, can weaken internal moral motivation and sometimes increase the tendency of good people to engage in misconduct (Alm, 2012; Ayala, Celse, & Hochman, 2019; Feldman, 2017; Pierce & Balasubramanian, 2015). For example, increasing the severity of punishment in the transport system led to higher fare evasion rates (Bijleveld, 2007). Additionally, employees may feel like they might be forgiven after reading a formal ethical code written in a more personal language (Kouchaki, Gino, & Feldman, 2019).

Insights from the growing behavioral ethics field suggest that people care for morality and may act unethically either unintentionally without being aware of their own acts or while knowingly breaking the rules while using self-justifications as a means to maintain a positive moral image (Bazerman & Gino, 2012). As a result, behavioral ethics has questioned the effectiveness of punishment and sanctions as ways to limit unethical behavior and suggested alternative methods to ensure honest behavior in organizations (Feldman et al., 2019). For example, the REVISE framework suggests three principles in shaping intervention: subtle cues to increase the salience of morality and decrease the ability to justify dishonesty, social monitoring cues to restrict anonymity, and self-engagement to increase people's motivation to maintain a positive self-perception as a moral person (Ayal, Gino, Barkan, & Ariely, 2015).

In addition, when referring to tax evasion, it seems that although enforcement may increase tax compliance, collecting taxes through aggressive enforcement and coercion is more costly than encouraging taxpayers to collaborate with tax collectors (Mascagni, 2018). As a result, there is a growing interest among policymakers and academics in improving "tax morale". Tax morale is an umbrella term that captures non-monetary motivations for tax compliance, such as intrinsic motivation to pay taxes or guilt for failure to comply, the influence of peer behavior, and cultural or social norms (Luttmer and Singhal 2014). Tax morale serves as a way to increase voluntary compliance with tax laws and ensure tax compliance even in situations where the ability to control and audit taxpayers is limited (Bott et al., 2019). For example, a study in a local church in Bavaria, where the German tax code legally obligates tax payment, found that even in an environment in which both actual and perceived enforcement of tax collection is absent, there is some degree of tax compliance due to intrinsic motivation (Dwenger, Kleven, Rasul, & Rincke, 2016).

Improving tax morale suggests that policymakers consider a rich set of instruments to affect compliance. For example, developing services to assist taxpayers in filing returns and paying taxes, using simple nudges to remind social norms, presenting information in a more accessible form, providing payment reminders, or using an advertisement that links taxes with government services (Alm, & Torgler, 2011; Lamberton, De Neve, & Norton, 2018; Luttmer, & Singhal, 2014). Importantly, this policy views the taxpayer as a client rather than a potential criminal and relies upon their own moral values and perception of the tax administration's quality and credibility (Alm, 2012).

Yet, despite the breadth of research on what drives unethical behavior and consequently policies and intervention, there is much less research on how unethical acts affect individuals'

subsequent behaviors toward the organization they have cheated. A better understanding of the downstream effects of unethical behavior may shed light on unexplored aspects of people's misbehavior. Thus, it may have far-reaching implications for organizations and policymakers in considering regulations and policies that mitigate unethical behavior and its effects.

Subsequent Judgment and Possible Impact on Policies

Unethical acts lead to significant financial losses but may also result in additional behavioral consequences that could affect the organization. Actors' subsequent behaviors may impact their satisfaction level from the organization, loyalty to the organization, and the nature of their (online) recommendations. For example, consider an employee that over-reported their monthly work hours. Would this unethical act influence the employee's judgment of the workplace? Would this over-reporting lead the employee to judge their workplace more or less positively? Or, consider a self-employed businessperson that underreported her annual income in order to reduce taxes owed. Would cheating influence her judgment and perception of the government authorities? Studies have shown that citizens' perceptions of government are based on their personal experience, judgments of the integrity and capability of public officials, and word of mouth (Houston, Aitalieva, Morelock, & Shults, 2016). Importantly, people's perceptions influence public trust, which is central to implementing public policies, and subsequently for effective, cooperative compliance (Gordon, 2000; Im, Cho, Porumbescu, & Park, 2014).

Critically, existing theories of unethical behavior provide conflicting predictions for how cheating may affect people's judgments of the victims of the unethical behavior. Because people wish to be moral and seen by others as such, unethical acts could increase feelings of guilt (Cohen, Wolf, Panter, & Insko, 2011; Shu & Gino, 2012). Guilt feelings could trigger reparative actions and prosocial behavior, such as empathy and care-giving, that could "balance the scale" (Gneezy, Imas, & Madarász, 2012; Tracy & Robins, 2007). According to this scenario, a cheating employee would judge and evaluate the workplace more favorably, attempting to relieve their guilt. However, according to some studies, compensation behavior may occur only if the person fails to resolve the ethical dissonance by justifying the act (Barkan, Ayal, Gino, & Ariely, 2012; Shalvi, Gino, Barkan, & Ayal, 2015). Unethical individuals may use justifications before or after engaging in dishonest acts and, as a result, will not regard their actions as unethical. In such cases, cheating will evoke less guilt and prosocial behavior (Hofmann, Wisneski, Brandt, & Skitka, 2014) and, therefore, might not influence their subsequent judgments.

However, people sometimes find it difficult to deny their misconduct and thus may still experience some ethical dissonance, a feeling of tension due to the inconsistency between their actual dishonest behavior and their desire to uphold a moral self-image (Ayal & Gino, 2011; Barkan et al., 2012). To alleviate that dissonance, people distance themselves from their misconducts, preferring to criticize others so they can present themselves as "ultra-moral" (Barkan et al., 2012; Shalvi et al., 2015). This account would suggest that people will judge their victim more negatively to preserve a sense of self-consistency with their moral history and past decisions (Jussim, Yen, & Aiello, 1995; Schwabe, Dose, & Walsh, 2018). The prediction that people would tend to judge the victim more negatively is also consistent with theories that suggest people can selectively disengage moral self-regulation (for instance, by attributing the blame to the victim) and behave unethically without feeling distressed (Bandura, 1999; Moore, Detert, Klebe Treviño, Baker, & Mayer, 2012). According to this account, the aforementioned cheating employee would judge and evaluate the workplace more negatively after cheating, trying to distance themselves from the object of their misconduct.

In summary, whereas theories that focus on guilt (e.g., Cohen et al., 2011) predict higher and positive judgments following dishonesty, theories of distancing (e.g., Barkan et al., 2012), self-consistency (Jussim et al., 1995), and self-regulation (Bandura, 1999) predict lower and negative judgments following dishonesty. Accordingly, I test these opposing directions among employees and taxpayers. Across four studies, I find that people tend to judge the organizations they have cheated more negatively, and the more they cheat, the more negative these judgments become. The results were found among employees who recalled their unethical behaviors in the workplace (Studies 1.1–1.3) and partially among self-employed who recalled their false tax reports (Study 1.4), implying that people tend to distance themselves by judging the target of their transgression in an attempt to justify their actions. Because unethical behaviors occurred in the past, I asked participants to recall their own ethical misconducts (Barkan et al., 2012). Previous studies showed the impact of recalling one's misconducts (compared to others') on subsequent moral self-image, intentions, and behavior (Jordan, Mullen, & Murnighan, 2011).

The experiments also demonstrate external validity in everyday, real-life situations among employees and taxpayers. As part of the increasing criticism of the behavioral ethics field for using abstract tasks in laboratory settings with low ecological validity (Ayal et al., 2019) and the tendency to shift research in behavioral ethics to a more prescriptive approach rather than a descriptive one (Ayal et al., 2016), I explored the impact of unethical behavior on

subsequent judgment among people in their daily life. The results enable a better understanding of the phenomenon and the possible implication for organizations and policymakers.

Additionally, the design of the studies enabled to dismiss a reversed causality account whereby a prior negative judgment allows people to cheat more. Most studies in the field of tax evasion emphasized that taxpayers' evasion levels depend on their satisfaction with public policy and their relationship with the authorities (Dell'Anno, 2009). However, using the recall paradigm (Fox & Kahneman, 1972), in which the satisfaction level from the organization was examined before or after recalling their transgression, showed a negative correlation between unethical behavior and subsequent judgment only after recalling their transgressions.

Unethical workplace behavior and judgment of the workplace (Studies 1.1-1.3)

The following studies examined employees' ratings of their employer as a function of the salience of unethical workplace behaviors. Unethical behavior was made salient using a recall paradigm (Blekher, Danziger, & Grinstein, 2019; Fox & Kahneman, 1992) that manipulated the order of two questions in a survey to examine whether the evaluation of satisfaction (or judgment in general) varies with the salience of another measure. In the studies, I asked employees to rate their workplace either before (low salience) or after (high salience) reporting whether they engaged in questionable behaviors in their workplace. I compared employees' ratings of their workplace in the two conditions to determine whether employees' judgments in the high salience condition would be more positive or negative than employees' judgments in the low salience condition.

Study 1.1 - Pretest

The pretest aim was to determine the number of questionable behaviors participants will be asked to recall. On the one hand, the number of questionable behaviors should be high enough to elicit thoughts and an in-depth examination of unethical behaviors in the workplace. On the other hand, the number of questionable behaviors should not be too high so as not to be seen as too difficult. In addition, I wanted to examine the perceived severity of each behavior to better understand the differences between the categories of the questionable behaviors (e.g., personal issues during work time, false reports).

Method

Participants. 196 participants (62% female, $M_{\text{age}} = 34$, $SD = 10.4$) who reported part- or full-time employment took part in the experiment. Participants were recruited from prolific,

online platform which offers a more diverse population and high data quality (Peer, Brandimarte, Samat, & Acquisti, 2017). Each participant was paid £0.40. The study used two conditions (2 or 12 questionable behaviors), with approximately 100 participants in each condition.

Design and procedure. Participants were asked to recall and write up to 2 or 12 questionable workplace behaviors ("Try to recall and write up to two/twelve questionable behaviors you have done personally in the last year in your workplace. There is no need to describe the context or to explain the behavior"). After recalling the questionable behaviors, participants were asked to indicate the severity of each act ("Please rate the severity of each behavior you mentioned on a scale of 1-7, 1 - not severe at all, 7 - very severe"). Then, participants were asked to indicate the extent to which they agree with two statements on a scale of 1-7 (1 – do not at all agree, 7 – agree completely): "it was easy for me to recall 2/12 questionable actions I've done in my workplace", "it would have been difficult for me to list more questionable acts than I did."

Results and Discussion

Participants who were asked to recall two questionable behaviors ($N = 101$) listed a significantly lower number of questionable behaviors compare to participants ($N = 95$) that were requested to recall 12 questionable behavior, $M = 1.95, 4.28$; $SD = .26, 3.16$; $t(194) = 7.39, p < .01$). However, the perceived severity of the questionable behaviors that were recalled among both groups was not significantly different ($M = 3.31, 3.02$; $SD = 1.37, 1.36$; $t(194) = 1.50, p = .14$).

There was no significant difference in the average ease of recalling questionable behaviors between participants that were requested to recall 2 vs. 12 actions ($M = 4.20, 3.91$; $SD = 1.95, 2.14$; $t(194) = 1, p = .32$). In addition, there was no significant difference in the average difficulty to recall additional questionable behaviors between participants that were requested to recall 2 vs. 12 actions ($M = 5.39, 5.72$; $SD = 1.63, 1.62$; $t(194) = -2.15, p = .16$), so that all participants perceived it as relatively difficult.

The most frequent categories of questionable behaviors participants wrote were: using company resources for personal benefit, using company time for personal issues, making false reports, and customer-related issues. However, questionable acts related to false reports (reporting extra hours) and customers (bad service) were perceived as the most severe questionable acts (Table 1.1).

Table 1.1. Perceived severity according to questionable act categories

Categories of questionable acts	Examples	Frequency	Perceived Severity
Using company resources for personal use	printing personal documents stealing stationery taking home company food or drinks	117	2.7
Personal issues during work time	doing personal tasks waste time online games	151	3.1
False reports	leaving early recorded extra hours pretend to be sick	72	4.1
Customer-related issues	ignore clients overcharging customers	20	4.8

According to the pretest results, more than 74% of participants provided more than merely two questionable behaviors when requested to recall up to 12 acts. Yet, only a few participants wrote more than ten questionable behaviors (10% of participants). Therefore, I concluded that participants would be asked to recall up to 10 questionable behaviors in the following studies.

Study 1.2 - Impact of perceived severity and prevalence of questionable behaviors on job satisfaction

In study 1.2, I used the recall paradigm (Blekher et al., 2019; Fox & Kahneman, 1992), which manipulated the order of two questions in a survey to examine whether the evaluation of satisfaction and judgment varies with the salience of another variable. In this study, I explore whether employees' subsequent workplace judgments vary with the salience of the perceived severity of questionable behaviors and their prevalence among their peers. Both measurements were an indication of how much participants perceived the questionable behaviors as normative and morally acceptable. I compared employees' ratings of their workplace in the two conditions to determine whether employees' judgments in the high salience condition would be more positive or negative than employees' judgments in the low salience condition.

Based on previous studies, unethical behaviors that are perceived as more severe may be considered less normative and threaten the moral self-image (Barkan et al., 2012). As a result, it may affect peoples' subsequent judgment of the target of the transgression. However, since current theories (guilt-relief and distancing) suggest opposing predictions on subsequent judgments, I predict differences in the satisfaction level between participants who cheated

compared to those who didn't cheat without a clear direction (supporting the guilt or distancing account):

H1.1 Employees who perceive their questionable behaviors as severe and recall it (high salience) will rate their satisfaction level significantly different compared to employees who perceive their questionable behavior as less severe and recall it, and compared to employees who report their satisfaction level before recalling their questionable behavior.

H1.2 There will be a significant correlation between the perceived severity of questionable behavior and the satisfaction level from the workplace only in the high salience condition.

Method

Participants. I recruited 280 participants from Prolific Academic (55% female, $M_{age} = 35$, $SD = 11.1$) who reported part- or full-time employment. All participants were above 18 years old, with English as their first language. Each participant was paid £0.40. The study used two conditions (low or high salience), with a minimum of 100 participants in each condition.

Design and Procedure. Participants performed two tasks. In one task, they were asked to recall and write up to 10 questionable workplace behaviors that they had engaged in over the last year. To ensure participants understood what we meant by questionable behaviors, I gave them three general examples of common questionable behaviors found in the pretest (using company resources for personal use, over-reporting to receive higher pay, and wasting time on non-work-related issues). Next, participants rated the severity of each behavior they mentioned on a scale of 1-7 (1 – not severe at all, 7 – very severe), and how often do they think other employees engage in each questionable act they mentioned on a scale of 1-5 (1 – never, 5 – a lot of the times). In the second task, participants completed a five-item job satisfaction survey (Andrews & Withey, 2012) on a scale ranging from 1 (extremely dissatisfied) to 7 (extremely satisfied). I used task order to make unethical behavior salient in participants' minds or not: participants rated their job satisfaction either before (low salience) or after (high salience) they reported their questionable workplace behaviors. In both conditions, the two surveys were presented as unrelated studies, and participants were asked to respond to each one separately.

Results and Discussion

There was no significant difference in the number of questionable behaviors participants in the high salience condition mentioned compared to those in the low salience

condition, $t(278) = .49$, $p = .63$. In addition, there was no significant difference in the overall workplace satisfaction rating between the conditions, $t(278) = .56$, $p = .57$. Importantly, there was no significant difference in the perceived severity between the conditions, $t(278) = .75$, $p = .46$, and on how often participants thought that other employees engaged in similar questionable behaviors, $t(278) = .40$, $p = .69$.

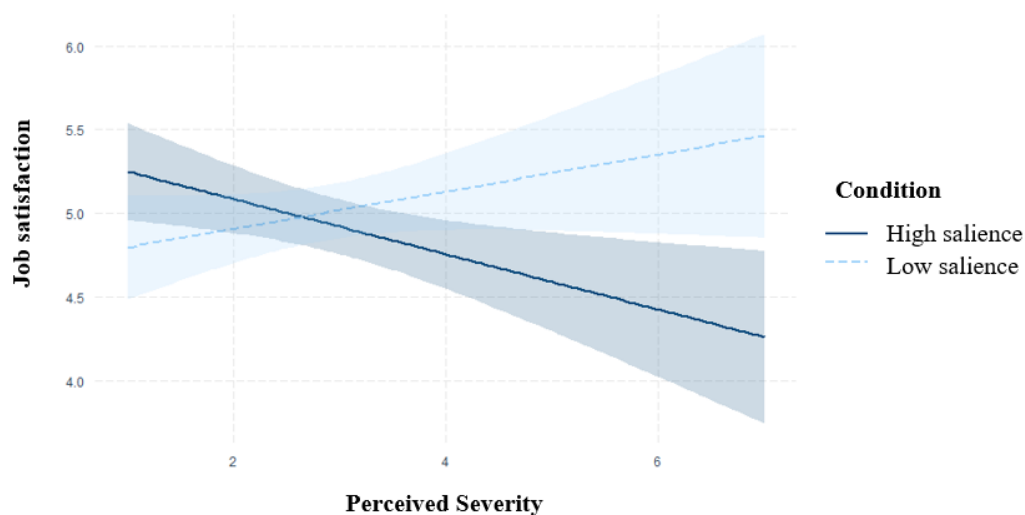
Table 1.2. Differences Between the Low and High Salience Condition

	Condition			
	Low Salience (N=140)		High Salience (N=140)	
	Mean	SD	Mean	SD
Number of questionable behaviors	3.44	2.30	3.59	2.50
Average Satisfaction	5.00	1.05	4.93	1.29
Perceived severity of the acts	2.87	1.15	2.97	1.35
Engagement of other employees in similar acts	3.62	0.84	3.66	0.94

To test hypothesis H1.1 (i.e., employees in the high salience condition who perceived their questionable behavior as more severe would report a significantly different satisfaction level compared to other participants), I conducted an ANOVA with the perceived severity (low versus high) and condition (low versus high salience) as independent variables and job satisfaction as the dependent variable. Because perceived severity is a continuous variable, I divided the perceived severity into high and low levels based on several different criteria: first, I tested differences between perceived severity that was lower/higher than the scale's midpoint (3.5 on the 1-7 scale). Second, I defined the low/high perceived severity based on the mean of the perceived severity and the median. The analysis revealed a significant interaction on the level of satisfaction, $F(1,276) = 4.89$, $p < .05$. Participants in the high salience condition who perceived their misconducts as severe reported the lowest job satisfaction ($M = 4.72$, $SD = 1.42$) compared to participants in the high salience condition who perceived their misconducts as less severe ($M = 5.03$, $SD = 1.20$), and compared to participants in the low salience condition who perceived their misconduct as less severe ($M = 4.90$, $SD = 1.09$) or more severe ($M = 5.25$; $SD = 0.91$). The results were significant also when I defined low and high perceived severity based on the mean ($M_{\text{perceived severity}} = 2.9$), $F(1,276) = 6.2$, $p < .05$, or based on the median ($M_{\text{perceived severity}} = 3$), $F(1,276) = 5.99$, $p < .05$.

Because the perceived severity is a continuous variable, and in order to examine hypothesis H1.2, I also tested the main effect of the correlation between perceived severity and job satisfaction in each condition. To test the hypotheses, I used PROCESS (Model 1 in Hayes, 2017) with the condition as the moderator for the effect of perceived severity of questionable behaviors on job satisfaction. The model was statistically significant, $F(3, 276) = 2.34, p < .05$, with a negative coefficient for the interaction ($b = -.28, SE = .11, p < .05$). As Figure 1.1 illustrates, there was a significant, negative correlation between the perceived severity of unethical behaviors and ratings in the high salience condition ($b = -.17, SE = .08, p < .05$), while in the low salience condition, the correlation was positive but not significant ($b = .12, SE = .08, p = .15$). These findings support hypothesis H1.2.

Figure 1.1. Impact of Perceived Severity of Questionable Behaviors on Job Satisfaction Rating in the Low and High Salience Conditions



The moderation model was not significant when considering the number of questionable behaviors participants mentioned on the level of satisfaction in both conditions ($F(3, 276) = .64, p = .59$), or the perceived prevalence of similar unethical behaviors among peers on the level of satisfaction ($F(3, 276) = .61, p = .60$). However, the moderation model was significant when considering both perceived severity of the questionable behavior and the prevalence of similar acts among peers (perceived severity by prevalence among peers) on the level of satisfaction from the workplace ($F(3, 276) = 2.68, p < .05$), with a negative coefficient for the interaction ($b = -.07, SE = .02, p < .05$). There was a significant, negative correlation between the perceived severity of unethical behaviors and satisfaction in the high salience condition ($b = -.20, SE = .08, p < .05$), while in the low salience condition, the correlation was positive but not significant ($b = .11, SE = .08, p = .22$).

These findings support the theories that predict that unethical behavior leads to a more negative subsequent judgment of the target. When employees recalled unethical behavior in the high salience condition, they subsequently judged their employer less favorably the more they perceived their behavior as more severe. Misbehaviors that are perceived as more severe may be considered less normative and elicit moral dissonance, an uncomfortable tension due to the inconsistency between moral self-image and the actual misbehavior (Barkan et al., 2012). Distancing from the wrongdoing by harsh judgment of the workplace can, therefore, serve as a post-violation justification to ease the tension and restore moral self-image.

Interestingly, the perceived prevalence of similar questionable behaviors among peers did not affect the workplace's subsequent judgment. However, when considering the perceived severity of the act with how prevalent it is, there was a negative impact on the workplace's satisfaction level. This finding may indicate that when a particular unethical behavior is perceived as common, it does not necessarily cause discomfort and may even be justified by "everyone is doing it" (Mazar, Amir, & Ariely, 2008). However, when this behavior is severe enough and perceived as relatively common among peers, it allows employees to blame the workplace more easily.

In addition to the importance of the perceived severity of the questionable behaviors, there is a need to consider the frequency in which employees engaged in certain misbehavior. Although there was no correlation between the number of questionable behaviors employees indicated in this study, the cumulative effect of a frequent engagement in a specific unethical action, even if not severe, can influence the ability to justify the behavior and hence on the subsequent judgment of the target of the transgression. Thus, the next study aims to replicate the findings from the current study (i.e., an employee who recalls their wrong did tend to judge their workplace less favorably) and extend it by reviewing the effect of the frequency of their wrongdoing.

Study 1.3 - Impact of engaging in questionable behaviors on job satisfaction

Study 1.3 uses the same recall paradigm as Study 1.2 (Blekher et al. 2019) to manipulate the salience of unethical behavior. However, in this study, I asked employees to rate their workplace either before (low salience) or after (high salience) they reported whether and to what extent they engaged in questionable behaviors in their workplace. Consistent with the distancing account, in the high salience condition, I expected a negative correlation between

the extent to which employees engaged in unethical behaviors and their workplace judgment, while in the low salience condition, I did not expect to find any significant correlation.

H1.3 Employees who cheat at a high frequency and recall it (high salience) will report the lowest job satisfaction level compared to employees who cheat at a lower level and recall it, and employees who report their satisfaction level before recalling their cheating.

H1.4 There will be a significant correlation between frequency of questionable behavior and the satisfaction level from the workplace only in the high salience condition.

Method

Participants. I recruited 272 participants from Prolific Academic (60% female, $M_{\text{age}} = 34$, $SD = 10.6$) who reported part- or full-time employment. Each participant was paid £0.40. The study used two conditions (low or high salience), with a minimum of 100 participants in each condition.

Design and Procedure. As in Study 1.2, participants performed two tasks. In one task, they were asked to recall and write up to 10 questionable workplace behaviors that they had engaged in over the last year. Next, participants indicated how often in the last year they had engaged in each of the behaviors they reported (1 = once, 2 = rarely, 3 = sometimes, 4 = frequently, 5 = almost every day). In the second task, participants completed a five-item job satisfaction survey (Andrews & Withey, 2012) on a scale ranging from 1 (extremely dissatisfied) to 7 (extremely satisfied). In this study, I also used task order to make unethical behavior salient in participants' minds: participants rated their job satisfaction either before (low salience) or after (high salience) they reported their questionable workplace behaviors. In both conditions, the two surveys were presented as unrelated studies, and participants were asked to respond to each one separately.

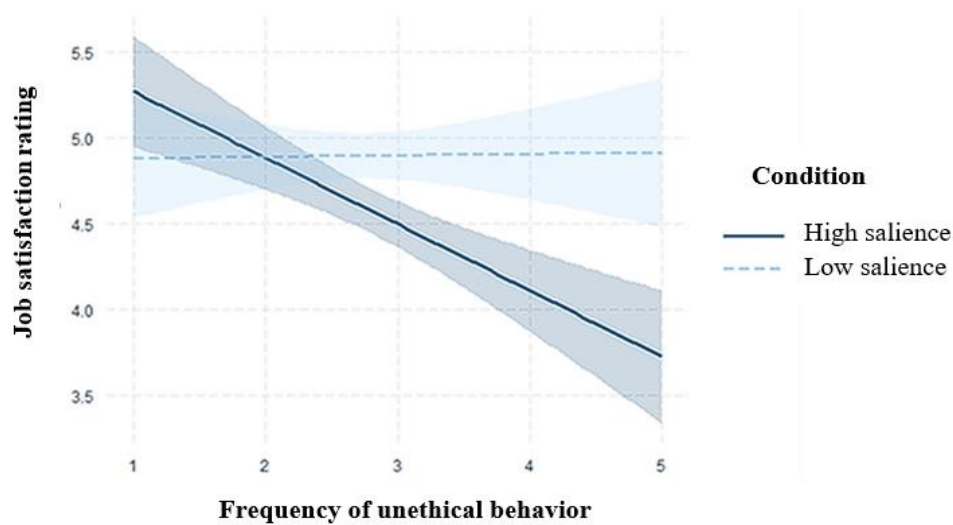
Results and Discussion

Participants in the high salience condition reported engaging in more questionable behaviors ($M = 4.65$, $SD = 2.86$) than those in the low salience condition ($M = 3.97$, $SD = 2.43$), $t(270) = 2.11$, $p < .05$, $\eta_p^2 = 0.016$. There was no significant difference in the average frequency of engagement in the reported behaviors between the high salience condition ($M = 2.79$, $SD = .79$) and the low salience condition ($M = 2.74$, $SD = .73$), $t(270) < .51$, $p = .61$. Job satisfaction ratings were lower in the high salience condition ($M = 4.58$, $SD = 1.30$) than in the low salience condition ($M = 4.90$, $SD = 1.10$), $t(270) = -2.15$, $p < .05$, $\eta_p^2 = 0.02$.

To test hypothesis H1.3 (i.e., employees who engage in questionable behavior and recall their behavior would report significantly different satisfaction level compared to other participants), I ran an ANOVA with the frequency of questionable behavior (low versus high) and condition (low versus high salience) as independent variables and job satisfaction as a dependent variable. Because the frequency of questionable behavior is a continuous variable, I divided the frequency into high and low levels based on several different criteria: first, I tested differences between frequency rate that is lower/higher than the scale's midpoint (2.5 on the 1-5 scale). Second, I defined the low/high frequency based on the mean frequency of questionable behaviors and the median. The analysis revealed a significant interaction on the level of satisfaction, $F(1,268) = 4.54, p < .05$. participants in the high salience condition who engaged in high frequency of questionable behaviors in their workplace reported the lowest job satisfaction ($M = 4.39, SD = 1.31$) compared to participants in the high salience condition who engaged in lower frequency in questionable behavior in their workplace ($M = 4.89, SD = 1.22$), and compared to participants in the low salience condition who engaged in low or high frequency in questionable behaviors in the workplace ($M = 4.81, 4.95; SD = 1.18, 1.05$). However, the results were not significant when I defined low and high frequency of questionable behavior based on the mean ($M_{\text{frequency}} = 2.76$), $F(1,268) = .48, p = .49$, or based on the median ($Md_{\text{frequency}} = 2.75$), $F(1,268) = .48, p = .49$.

Because the frequency of questionable behavior is a continuous variable, and in order to examine hypothesis H1.4, I also tested the main effect of the correlation between the frequency of questionable behavior and job satisfaction in each condition. To test the hypotheses, I used PROCESS (Model 1 in Hayes, 2017) with the condition as a moderator for the effect of frequency of engagement in questionable behaviors on job satisfaction. I used the number of questionable behaviors participants wrote (from 1 to 10) as a covariate to control for possible effects of verbosity (in writing skills) or recency (e.g., people who acted unethically more recently could have recalled more instances). The model was statistically significant, $F(4, 267) = 3.64, p < .01$, with a negative term for the interaction, $b = -.41, SE = .19, p < .05$. As Figure 1.2 illustrates, I found, consistent with hypothesis H1.4, a significant, *negative* correlation between frequency of unethical behaviors and ratings in the high salience condition, $b = -.24, SE = .14, p < .01$, while in the low salience condition, the correlation was *not* significant, $b = .01, SE = .13, p = .92$.

Figure 1.2. Impact of engaging in questionable behaviors (frequency of unethical behaviors in the last year) on job satisfaction rating in the high salience and low salience conditions



Although not part of the main research hypotheses, I also examined whether the employee's role (i.e., not a manager, first-level manager, or senior manager) is significant to the correlation between frequency of unethical behavior and job satisfaction. The analysis can be interesting since previous studies found that managers may be negatively affected by challenging goal settings and tend to use questionable tactics such as realizing sales revenue early or hiding expenses to get rewarded while indicating that the goal itself was the source of the problem (Ordóñez & Welsh, 2015). The analysis revealed that the employee's role is significant only among first-level managers (supervisors) in the high salience condition. In the high salience condition, there was a negative correlation between the frequency of first-level manager's unethical behaviors and job satisfaction rating ($b = -.42$, $F(1,80) = 16.9$, $p < .05$).

The findings of the study support the theories that predict that unethical behavior leads to more negative subsequent judgments of the target of the transgression. When employees recalled their unethical behavior in the high salience condition, they subsequently judged their employer less favorably the more they had recalled such behaviors. This pattern supports the distancing account, in which people distance themselves from their wrongdoing so they can view themselves as moral (Barkan et al., 2012). Considering that in 2019, one out of every 50 employees was caught stealing from their employer (Hayes international annual retail theft survey, 2019), it may have a far-reaching implication on the level of satisfaction of employees,

the word of mouth of the employees, and thus the ability to recruit and maintain good employees in the organization.

Paradoxically, these findings suggest that, *ceteris paribus*, the more lenient a workplace is in enabling and allowing (or not limiting or prosecuting) unethical workplace behaviors, the less satisfied its workers may be. However, it can also support interventions that increase the salience of moral rules and previous behaviors to decrease unethical behavior. In addition, it supports the findings that challenging goals may have a negative effect on the managers and may encourage unethical behaviors (Ordóñez & Welsh, 2015).

One may argue that finding a significant negative correlation between frequency of unethical behaviors and ratings in the high salience condition does not test causality. Meaning, it may be that when employees are less satisfied with their workplace, they tend to engage more frequently in unethical behaviors and not the opposite. However, if these were (only) correlational findings, I would have expected to find the same, or a similar, correlation between unethical behavior and employee ratings in the low salience condition, which I did not. The next study was designed to further test these findings' replicability in another sample and a different organizational setting.

Study 1.4 - Tax evasion and judgment of tax authorities

In Study 1.4, I focused on false tax reporting, one of the most common ethical violations among individuals (Lamberton, De Neve, & Norton, 2018). More specifically, I explored the impact of false tax reporting of the self-employed on their ratings of relevant government tax authorities, using the same recall manipulation used in Studies 1.2 and 1.3. Departing from them, I asked participants to report the frequency with which they engaged in various types of false tax reporting, among other behaviors not related to cheating. Consistent with the findings of Studies 1.2 and 1.3, I hypothesized that participants in the high salience condition would selectively rate the government agencies that deal with tax and business related issues more negatively the more they report unethical tax reporting related to those agencies.

H1.5 Self-employed who cheat at a high level and recall it (high salience) will report the lowest satisfaction level from business and tax-related authorities compared to participants who recall their actions and cheat to a small extent, and compared to participants who report their satisfaction level before recalling their questionable actions (whether the cheating level is high or low).

H1.6 There will be a significant negative correlation between unethical tax reporting and the rating of government agencies only in the high salience condition, and only when judging government agencies related to business or tax.

Method

Participants. I recruited 309 workers (54% female, $M_{\text{age}} = 41$, $SD = 11.8$) that were all pre-screened to be self-employed, using a commercial online panel in Israel. Each participant received 20 points that can be converted into gift cards (equivalent to 2 NIS, about \$0.60) for completing the study. The study used two conditions (low or high salience), with a minimum of 100 participants in each condition.

Design and Procedure. Participants in the high salience condition first indicated how often they perform ten behavioral items that indicate tax evasion (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always) and six business-related behaviors unrelated to taxes, all presented together in random order. Items related to tax evasion were: I pay in cash (without an invoice); I report my family cleaning products' expenses as business expenses; I report my family fuel expenses as business expenses; I report my family flights expenses as business expenses; I report my family grocery expenses as business expenses. Items related to honest tax reporting (reverse-coded) were: I report on each of the jobs I perform in my business; I issue invoices for bank transfers; I issue invoices at the actual payment date; I issue an invoice for every cash payment; I report the full amount of cash payments. The items unrelated to taxes were: I'm available to clients on the phone; I work on weekends and holidays; I work especially from home; I regularly update my customer base; I'm active in social media; I'm available to customers also during the evening.

Then, participants rated their satisfaction (from 1 = not at all satisfied, to 5 = very satisfied) with five government authorities that deal with taxes and businesses, and 13 authorities that do not. Agencies related to tax reporting and businesses were determined in advance and included: Tax Authority, VAT Authority, Social Security, Small Business Agency, Registrar of Companies. The unrelated agencies were: Ministry of Education, Ministry of Health, Ministry of Culture, Ministry of Environmental Protection, Fire Department, Ministry of Tourism, Consumer Protection Authority, Standards Institute, Ministry of Social Equality, Police, Ministry of Energy, Immigration Authority, Ministry of Communication.

In the low salience condition, participants first rated their satisfaction from government agencies and then reported the frequency of behaviors related and unrelated to tax evasion. In the high salience condition, participants first reported the frequency of behaviors related to tax reporting and only then rated their satisfaction level from government authorities. Lastly, participants reported their demographics and were thanked for their participation.

Results and Discussion

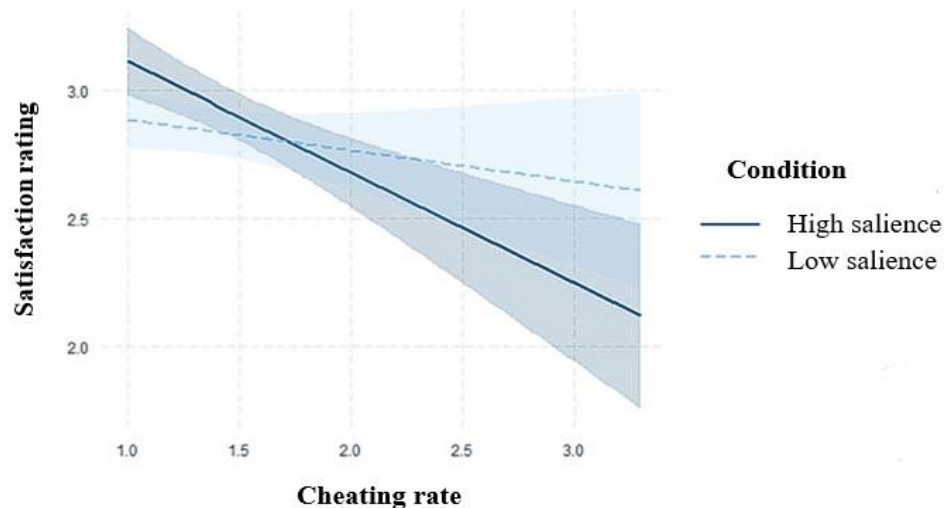
I found that participants in the low salience condition reported less tax evasion behaviors ($M = 1.39$, $SD = .45$) compared to participants in the high salience condition ($M = 1.51$; $SD = .50$), $t(306) = 2.15$, $p < .05$, $\eta_p^2 = 0.014$. In addition, there was no statistically significant difference in satisfaction ratings between the low salience condition ($M = 2.74$, $SD = .70$) compared to the high salience condition ($M = 2.81$; $SD = .81$), $t(306) = -0.75$, $p = .45$, $\eta_p^2 = 0.002$.

To test hypothesis H1.5 (i.e., self-employed would be less satisfied with business and tax-relevant authorities when their tax evasion behaviors were high and made salient), I ran an ANOVA with the cheating rate (low versus high) and condition (low versus high salience) as independent variables and satisfaction rating as a dependent variable. Because the cheating rate is a continuous variable, I divided the cheating into high and low levels based on several different criteria: first, I tested differences between cheating rate that is lower/higher than the scale's midpoint (2.5 on the 1-5 scale). Second, I defined the low/high cheating rate based on the mean of the cheating rate and the median. The analysis of the interaction on the level of satisfaction was significant only when low and high cheating rated were defined based on the median ($M_{\text{cheating rate}} = 1.30$), $F(1,305) = 6.61$, $p < .05$. However, the interaction on the level of satisfaction was not significant based on the scale's midpoint, $F(1,305) = .13$, $p = .71$, or based on the mean ($M_{\text{cheating rate}} = 1.45$), $F(1,305) = .38$, $p = .54$.

Because the cheating rate is a continuous variable, and in order to examine hypothesis H1.6, I also tested the main effect of the correlation between perceived severity and job satisfaction in each condition. To test the hypotheses, I used PROCESS (Model 1 in Hayes, 2017) with task order as a moderator for the effect of tax evasion on satisfaction from government authorities. The model was statistically significant, $F(3, 305) = 3.11$, $p < .05$, when using the combined ratings of the relevant authorities (e.g., tax authorities, VAT authorities, social security, small business agency, registrar of companies; Cronbach's $\alpha = .80$). However, the interaction was not significant ($b = -.31$, $SE = .21$, $p = .14$). As figure 1.3 illustrates, there

was a significant and negative correlation between the extent of engagement in tax evasion and satisfaction rating in the high salience condition, $b = -.22$, $SE = .16$, $p < .01$, while in the low salience condition, the correlation was not significant, $b = -.07$, $SE = .14$, $p = .38$.

Figure 1.3. Effect of tax evasion on satisfaction rating of government authorities



Results were significant also when considering only authorities that deal directly with taxes (i.e., tax authority, VAT authority, social security). A PROCESS analysis (Model 1 in Hayes, 2017) with task order as a moderator for the effect of tax evasion on satisfaction from government authorities, was statistically significant, $F(3, 305) = 3.11$, $p < .05$, with a non-significant interaction, $b = -.32$, $SE = .25$, $p = .21$. The correlation between cheating level and satisfaction rating in the high salience condition was significant $b = -.22$, $SE = .19$, $p < .05$, while in the low salience condition, the correlation was not significant, $b = -.08$, $SE = .17$, $p = .29$.

When participants first recalled their unethical behavior and only then indicated their satisfaction with different authorities (high salience condition), they expressed lower ratings for the tax-related authorities the more they reported having engaged in tax evasion. Importantly, consistent with hypothesis H1.6, the effect was significant only for tax-relevant authorities, and not for tax-unrelated authorities, $F(3, 305) = .99$, $p = .40$.

Considered jointly, the results of the study are mixed and do not fully support the hypotheses. Hypothesis H1.5 is not supported, and the interaction of the moderation model (Figure 1.3) was not significant. A possible explanation for these mixed findings is that the average reported cheating rate was relatively low ($M = 1.45$), with 36% of participants reporting higher than average cheating rate and only a few participants reporting a cheating

rate higher than the scale's midpoint (5.6% of participants). Future studies can use a larger sample to achieve more conclusive results. In addition, since taxpayers may be less inclined to report their tax evasion, it may be useful to ask participants to recall and write their tax-related behaviors (similar to studies 1.2-1.3) instead of using a list of behaviors that may be perceived as illegal.

Although not fully supported, it is still evident that an increase in the extent and saliency of cheating has a negative effect on subsequent evaluations and judgments of the organization towards which the transgression was targeted. More specifically, self-employed workers rated tax-related authorities more negatively the more salient (and frequently reported) their tax-evasion acts were. It appears that participants attempted to distance themselves from their unethical actions of tax evasion, presumably to justify their transgressions, thereby decreasing their ethical dissonance and preserving their moral self-image (Mazar et al., 2008). Interestingly, a negative correlation between tax evasion and satisfaction level was relevant only to tax-related authorities, implying that negative judgment is limited to those that were cheated. In addition, the results challenge the common notion that people engage in tax evasion due to low satisfaction levels with public policy and the quality of their relationship with the authorities (Dell'Anno, 2009). If this was the case, the low salience condition results would suggest a similar negative correlation between the satisfaction level of self-employed and tax evasion. In addition, we would expect to find a negative correlation between tax evasion to other governmental authorities, expressing a low satisfaction level from public policy in general and not necessarily only with tax-related authorities.

General Discussion

The research sheds light on an important and understudied phenomenon: how people judge organizations they have cheated. In four experiments, I show that when people behave unethically, they tend to judge the victim more harshly, presumably to distance themselves, and thus justify their transgression. Employees rated their workplace less favorably when the perceived severity of their unethical behavior or the frequency of their misbehaviors increased (hypotheses H1.1-H1.2 and H1.3-H1.4). However, there was no significant correlation between the perceived prevalence of unethical behavior among peers and satisfaction level. Although it is possible that participants found relatively common behaviors easier to justify, and therefore did not feel the need to use post-justifications such as distancing account (Shalvi et al., 2015), future studies might explore this aspect to better understand the boundary conditions of the phenomenon. Research might also examine whether informing people of the desire social

norms and what others are doing (i.e., descriptive norms; Ayal et al., 2019), not only relying on their own perception, will influence the subsequent judgment of those they have cheated.

In addition, I found a significant negative correlation between cheating and ratings only in the high salience condition, thus refuting a reverse-causality argument whereby people judge the victim negatively and therefore are more willing to cheat. The results were consistent in both organizational settings of employees toward their workplace and self-employed toward tax-related authorities (hypothesis H1.6). However, the findings did not fully support hypothesis H1.5. The satisfaction level of self-employed with a high cheating level in the high salience condition was significantly different from other participants, only when the median of cheating determined the level of cheating (low or high). The findings may suggest that contrary to previous studies (e.g., Dell'Anno, 2009) and common beliefs among the general public, it is not necessarily the low satisfaction from public policy that drives tax evasion but rather the opposite. Meaning, for some people, it is possible that tax evasion impacts their low satisfaction from public policy and trust in tax authorities. However, due to the inconsistency in findings, further studies are needed to address the issue.

The research adds to the extant literature by considering not only the direct costs of unethical behavior (e.g., lost revenues) but also its lingering, less direct effects on judgments and word of mouth in long-term relationships, such as with an employer or the government. By revealing these lingering effects, the research underscores the importance of developing effective policies that mitigate unethical behavior and its effects. For example, my findings imply that policies directed to increase trust and satisfaction among taxpayers and employees have to consider the impact of organizations' unethical behaviors. Also, organizations need to consider a setting that increases the salience of morality and previous behaviors to discourage people from cheating and sever the tie between cheating and subsequent negative judgments (Ayal et al., 2015).

Furthermore, the research findings have important implications for organizations because of the systematic effects that a person's cheating may have on how they communicate with others about the organization they cheated. In today's digital environment, organizations are subject to public opinion in the form of online reviews, blogs, social media, and internet forums (Chuang, 2020), and a cheating individual who decides to harm the victim of their transgression can do so. My findings suggest that these communications may be negatively biased because unethical individuals may distance themselves from their acts to restore their

moral self-image. However, the extent of this phenomenon is unclear because unethical people may choose not to communicate their opinions about the victims. Future studies might examine this conjecture to better understand the direct and indirect impacts of unethical behavior on online and offline person-generated communication. A particularly fruitful avenue may be to explore the link between employee dishonesty, the employee's communications regarding their employer, and recipients' opinion of that employer.

Limitations and future research directions

The research contributes to the knowledge on the consequences of unethical behavior by being the first to examine its effect on subsequent judgment of the victim. Although my findings suggest that people tend to distance themselves by judging their victims harshly, further studies are needed to reinforce the findings. In addition, a better understanding of the underlying mechanism driving the effect is still needed. For example, manipulating the ability to justify the unethical behavior may support the distancing-based process by revealing whether other mechanisms, such as guilt, are being evoked when it is not possible to justify the act and blame the target of their transgression. I will address the potential moderating role of the ability to justify misbehavior in chapter 2.

In addition, this research explored individuals' subsequent behavior of the victim in long-term interactions that characterize organizations such as workplaces and governmental entities. However, future research may examine whether the pattern still holds in one-time interactions, such as when using a product or service. Exploring the impact of one-time interactions in the marketing realm can be especially insightful since cheating can impact a brand's ability to transform one-time interaction into an affective commitment and behavioral loyalty (Steenhaut & Van Kenhove, 2005). Moreover, previous studies suggest that unethical behaviors can be examined by whether people are taking active or passive advantage of the act and by the existence of a salient victim (Muncy & Vitell, 1992). Future research may aim to uncover these factors and the moderating effect they may have on subsequent judgment of the victim.

Conclusion

In conclusion, this research uncovers an unexplored consequence of unethical behavior, which is how people judge those they have cheated. Behaving unethically may encourage people to distance themselves and harshly judge their victims, providing lower ratings and potentially causing further harm to those they have already cheated. This research provides important

implications for those responsible for maintaining an organization's image in the minds of the employees and the general public. It also has potential implications for policy in general: lenient policies that put less focus on preventing or enforcing unethical behavior, perhaps with the intention of conferring more trust on the regulated parties, actually expose themselves to a paradoxical effect – the more unethical behaviors they allow, the more unethical behaviors they invite, which can then lead to less favorable evaluations of the organization.

References

- ACFE (2020). 2020 *Global study on occupational fraud and abuse*. Retrieved from <https://acfepublic.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf>
- Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International tax and public finance*, 19(1), 54-77.
- Alm, J., & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635-651.
- Andrews, FM., & Withey, SB. (2012) *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* (Springer, Boston).
- Ayal, S., Celse, J., & Hochman, G. (2019). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes*.
- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. Mikulincer M, Shaver PR, eds. *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (American Psychological Association, Washington, DC) 149-166.
- Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science*, 10(6), 738-741.
- Ayal, S., Hochman, G., & Ariely, D. (2016). Dishonest behavior, from theory to practice. *Frontiers in Psychology*, 7, 1521.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757-773.
- Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, 8, 85-104.
- Bijleveld, C. (2007). Fare dodging and the strong arm of the law. *Journal of experimental criminology*, 3(2), 183-199.

- Blekher, M., Danziger, S., & Grinstein, A. (2019). Salient volunteering behavior increases monetary risk-taking. *Journal of Consumer Psychology*. DOI: 10.1002/jcpy.1146
- Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2019). You've got mail: A randomized field experiment on tax evasion. *Management Science*.
- Business Software Alliance (2018). *2018 BSA Global Software Piracy Study*. Retrieved from <https://gss.bsa.org/>.
- Chuang, S. H. (2020). Co-creating social media agility to build strong customer-firm relationships. *Industrial Marketing Management*, 84, 202-211.
- Cohen, TR., Wolf, ST., Panter, AT., & Insko, CA. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, 100(5), 947-966.
- Dell'Anno, R. (2009). Tax evasion, tax morale and policy maker's effectiveness. *The Journal of Socio-Economics*, 38(6), 988-997.
- Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203-32.
- Feldman, Y. (2017). Using Behavioral Ethics to Reduce Organizational Misconduct. *Forthcoming, Behavioral Science and Policy Special Volume on Corruption*.
- Feldman, Y. (2018). *The law of good people: Challenging states' ability to regulate human behavior*. Cambridge University Press.
- Feldman, Y., van Rooij, B., & Rorie, M. (2019). Rule-breaking without Crime: Insights from Behavioral Ethics for the Study of Everyday Deviancy. *Feldman, Y., Rorie, M., & Van Rooij, B.(2019). Rule-breaking without Crime: Insights from Behavioral Ethics for the Study of Everyday Deviancy. The Criminologist*, 44(2), 8-11.
- Fox, C. R., & Kahneman, D. (1992). Correlations, causes and heuristics in surveys of life satisfaction. *Social Indicators Research*, 27(3), 221-234.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1-44.

- Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current opinion in behavioral sciences*, 3, 107-111.
- Gill, D., Prowse, V., & Vlassopoulos, M. (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization*, 96, 120-134.
- Gneezy, U., Imas, A., & Madarász, K. (2012). Conscience accounting: Emotional dynamics and social behaviour. *Theoretical economics paper series* (STICERD, London).
- Gordon, M. T. (2000). Public trust in government: The US media as an agent of accountability?. *International Review of Administrative Sciences*, 66(2), 297-310.
- Hayes, AF. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (Guilford Press, New York).
- Hayes international annual retail theft survey (2019). *32nd Annual retail theft survey*. Retrieved from <http://hayesinternational.com/news/annual-retail-theft-survey>
- HMRC (2019). *Overview of HMRC's annual report and accounts 2018 to 2019*. Retrieved from <https://www.gov.uk/government/publications/hmrc-annual-report-and-accounts-2018-to-2019/overview-of-hmrCs-annual-report-and-accounts-2018-to-2019>
- Hofmann, W., Wisneski, DC., Brandt, MJ., & Skitka, LJ. (2014). Morality in everyday life. *Science*, 345(6202), 1340-1343.
- Houston, D. J., Aitalieva, N. R., Morelock, A. L., & Shults, C. A. (2016). Citizen trust in civil servants: A cross-national examination. *International Journal of Public Administration*, 39(14), 1203-1214.
- Im, T., Cho, W., Porumbescu, G., & Park, J. (2014). Internet, trust in government, and citizen compliance. *Journal of Public Administration Research and Theory*, 24(3), 741-763.
- IRS - Internal Revenue Service (2016). *Tax Gap Estimates for Tax Years 2008-2010*. Retrieved from <https://www.irs.gov/newsroom/the-tax-gap>
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, 37(5), 701-713.

- Jussim, L., Yen, H., & Aiello, JR. (1995). Self-consistency, self-enhancement, and accuracy in reactions to feedback. *Journal of Experimental Social Psychology*, 31(4), 322-356.
- Kouchaki, M., Gino, F., & Feldman, Y. (2019). The ethical perils of personal, communal relations: A language perspective. *Psychological science*, 30(12), 1745-1766.
- Lamberton, C., De Neve, J. E., & Norton, M. I. (2018). The power of voice in stimulating morality: Eliciting taxpayer preferences increases tax compliance. *Journal of Consumer Psychology*, 28(2), 310-328.
- Luttmer, E. F., & Singhal, M. (2014). Tax morale. *Journal of economic perspectives*, 28(4), 149-68.
- Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, 32(2), 273-301.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6), 633-644.
- Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, 65(1), 1-48.
- Mulder, L. B. (2018). When sanctions convey moral norms. *European Journal of Law and Economics*, 46(3), 331-342.
- Muncy, J. A., & Vitell, S. J. (1992). Consumer ethics: An investigation of the ethical beliefs of the final consumer. *Journal of business Research*, 24(4), 297-311.
- National Retail Federation (2020). *Consumer Returns in the Retail Industry 2020*. Retrieved from <https://nrf.com/research/customer-returns-retail-industry>
- Ordóñez, L. D., & Welsh, D. T. (2015). Immoral goals: how goal setting may lead to unethical behavior. *Current Opinion in Psychology*, 6, 93-96.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153-163.
- Pierce, L., & Balasubramanian, P. (2015). Behavioral field evidence on psychological and social factors in dishonesty and misconduct. *Current Opinion in Psychology*, 6, 70-76.

- Schwabe, M., Dose, D. B., & Walsh, G. (2018). Every Saint has a Past, and Every Sinner has a Future: Influences of Regulatory Focus on Consumers' Moral Self-Regulation. *Journal of Consumer Psychology*, 28(2), 234-252.
- Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, 23(10), 1264-1270.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125-130.
- Shu, LL., & Gino, F. (2012). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*, 102(6), 1164-1177.
- Singh, P., & Twalo, T. (2015). Mismanaging unethical behaviour in the workplace. *Journal of Applied Business Research (JABR)*, 31(2), 515-530.
- Steenhaut, S., & Van Kenhove, P. (2005). Relationship commitment and ethical consumer behavior in a retail setting: The case of receiving too much change at the checkout. *Journal of Business Ethics*, 56(4), 335-353.
- Taylor, E. (2016). Supermarket self-checkouts and retail theft: The curious case of the SWIPERS. *Criminology & Criminal Justice*, 16(5), 552-567.
- Tracy, JL., & Robins, RW. (2007). The self in self-conscious emotions: A cognitive appraisal approach. Tracy JL, Robins RW, Tangney JP, eds. *The Self-Conscious Emotions: Theory and Research* (Guilford Press, New York), 3-20.
- Weber, J., Kurke, L. B., & Pentico, D. W. (2003). Why do employees steal? Assessing differences in ethical and unethical employee behavior using ethical work climates. *Business & Society*, 42(3), 359-380.
- Zhang, T., Gino, F., & Bazerman, M. H. (2014). Morality rebooted: Exploring simple fixes to our moral bugs. *Research in Organizational Behavior*, 34, 63-79.

Chapter 2

Consumer's Unethical Behavior, Self-Justification, and Subsequent Judgments

Abstract

How do consumers judge those they have cheated? Brands and retailers aim to build a strong consumer-brand relationship to gain loyal customers who serve as brand advocates and recommend it to others. However, although previous studies suggest that many customers behave unethically, it is unclear whether these misbehaviors impact their subsequent judgments of the brand and the possible underlying mechanism that drives the effect. In five studies, I test the impact of unethical behavior in common consumption situations and examine possible factors that moderate it. Study 2.1 provides a partial support for the distancing account, in which consumers tend to harshly judge those they have cheated. However, the tendency to harshly judge the victim of their transgression is only relevant when people can justify the cheating (Study 2.2). In addition, I find support for the moderating role of moral credentials (Study 2.3) and how consumers tend to increase their profit when passively benefit from cheating (Studies 2.4 and 2.5). These findings carry important managerial implications because they impact word of mouth and reveal boundary conditions to the phenomenon in general.

Introduction

How do consumers judge those they have cheated? Consider a shopper who uses a self-checkout cashier at the supermarket and shoplifts by not scanning an item. Will this unethical act influence the shopper's judgment of the supermarket? When asked to rate the supermarket, will having "cheated" impact the shopper's evaluation? While much research has focused on the reasons, antecedents, and remedies for unethical consumer behavior (e.g., Ayal, Hochman & Ariely, 2016; Gerlach, Teodorescu, & Hertwig, 2019), I examine an important, yet understudied consequence of dishonesty – consumers' judgment and reviews of sellers, vendors, or organizations they have cheated. Given that 64% of consumers are estimated to have engaged in unethical consumption at some point in their lives (Farmer & Dawson, 2017) and because consumers' judgments drive brand perception and online reviews, understanding whether unethical behavior has systematic subsequent influences is important.

While many interactions in the marketing realm could be one-time incidents, vendors and sellers aim to develop relationships with the customers and transform those interactions into affective loyalty (Steenhaut & Van Kenhove, 2005). However, due to the unique characteristics of one-time interactions (e.g., lack of commitment, the vendor's anonymity; Gino, Shu, & Bazerman, 2010), the impact of unethical behavior on subsequent judgments and behaviors may differ from other contexts. Moreover, while the behavioral ethics field considers self-justification as a way to mitigate the threat to the moral self-image (Shalvi, Gino, Barkan, & Ayal, 2015), its moderating role in the impact of unethical behavior on subsequent judgments is still unclear. The ease with which people can justify their cheating may have a significant moderating effect and thus implications to business and marketing strategies, as well as for policymakers.

Unethical Consumption

Unethical consumption includes acts that "violate the generally accepted norms of conduct in consumption situations, and thus disrupt the consumption order" (Fullerton & Punj, 2004, p. 1239). These acts include a myriad of behavior in an online and offline environment, ranging from lying about a child's age to receive a lower price, through not saying anything when receiving too much change, all the way to shoplifting and insurance frauds (Vitell & Muncy, 2005). Although many mundane ethical violations are small enough so that consumers can easily justify them to themselves (Shalvi, Eldar & Bereby-Meyer, 2012), they nonetheless account for billions of dollars in lost revenues annually. For example, recent data suggests that

"wardrobing" (using clothes once, then returning them for a full refund) costs the U.S. retail industry an estimated \$10.2 billion annually (National Retail Federation, 2020), shoplifting and employee theft account for \$42 billion lost in retail revenue (Taylor, 2016), with an estimated loss of £246 million to convenience stores in 2018 due to unethical acts of customers (Association of Convenience Stores, 2019). In addition, illegal software downloads totaled \$46.3 billion globally in 2018 alone (Business Software Alliance, 2018).

According to Muncy & Vitell (1992), one way to classify dishonest actions in the marketing realm is by evaluating them based on two variables – taking active or passive advantage of the action and the existence of a salient victim. As a result, unethical practices can be divided into four categories: The first one, 'actively benefiting from illegal activity', comprises actions that are initiated by the consumer and are almost universally perceived as illegal (e.g., shoplifting). The second, 'passively benefiting at the expense of others', involves taking advantage of a seller's mistake (e.g., receiving too much change and not saying anything). The third, 'actively benefiting from a questionable action', includes actions that may not necessarily be perceived as illegal but still harm the seller in some way (e.g., accidentally damaging an item and not saying anything). The last one, 'no harm/no foul', includes actions that do little or no harm (e.g., trying on clothes for two hours and not buying anything eventually, or working for few hours in a not too crowded coffee shop over a small cup of coffee).

Most of the everyday questionable acts from consumer and marketing domains fit in the second and third categories and may sometimes refer to small actions in the first category (i.e., shoplifting, small theft). These acts represent the “many apples in the barrel that turned just a little bit bad” (Ayal & Gino, 2011; p. 3) and force retailers to face the increasing costs of lost merchandise while also allocating more budget in favor of expensive measures for prevention, detection, and prosecution of such acts (Yaniv, 2009). For example, in 2019, 44% of retailers in the US allocated additional technology resources and increased their loss prevention budget (NRF, 2019). Preventative measures include technology-based solutions such as facial recognition or RFID tags, intervention by security staff (Potdar, Guthrie, Gnoth, & Garry, 2018), and store design to signal consumers about norms and deter them from unethical behavior (e.g., using signs, symbols, lights, and specific colors) (Fombelle et al. 2020).

However, according to recent data, consumer misbehavior is on the rise (PriceWaterhouseCooper, 2018), and the effectiveness of current measures is increasingly being questioned. Thus, there is a growing interest in addressing alternative avenues of unethical consumption prevention while considering the antecedents of unethical consumption and the target of the transgression (Potdar et al., 2018). Yet, while research has focused on the scope and causes of unethical behavior (Ayal et al., 2016; Gerlach et al., 2019; Gino & Ariely, 2016), a better understanding of the impact of unethical behavior on subsequent judgments may have implications in terms of policies and marketing strategies. For example, a better understanding of the phenomenon may emphasize the need to shift resources from expensive preventive measures to more creative marketing efforts along the customer journey.

Subsequent Judgment in Marketing Realm

Consumers' subsequent judgments and behaviors toward the firm they have harmed may have a significant impact in the marketing realm. For example, consumer satisfaction levels from a product or a seller may differ depending on whether they purchased or consumed the product/service ethically or not. Similarly, consumers' loyalty to a brand, the nature of reviews consumers may post online, their inclination to repurchase the same product or brand, and other related behaviors may also be affected by whether they consumed the product ethically or not. Consider, for example, the consumer that shoplifted by not scanning an item in the self-checkout cashier at the supermarket. It is unclear whether that consumer (compared to one who did not engage in any unethical consumption) would exhibit a different level of satisfaction from the supermarket, or whether they would be more or less likely to provide an online review for that supermarket and repurchase from it again.

Most of the studies that have explored possible consequences of unethical consumption focused mainly on the effects of unethical behavior on post-cheating emotional reactions (e.g., Ruedy, Moore, Gino, & Schweitzer, 2013), the propensity for prosocial activities towards other entities, such as donations to charity or volunteering (Gneezy, Imas & Madarasz, 2012), and the impact of unethical behavior on guilt and guilt-relief (Cohen, Wolf, Panter & Insko, 2011; Xu, Be'gue & Bushman, 2012). One study explored the phenomenon of emotional accounting, in which consumers who possessed "negatively tagged money" (i.e., obtained illegally) preferred to use it for more virtuous causes, such as to fund someone's education (Levav & McGraw, 2009). Critically, however, these studies all focused on the effects of unethical behavior on people's subsequent behavior *outside* the context of their transgression, and

typically do not consider how unethical consumption can impact consumers' judgments and behaviors towards the very entity (brand, vendor, or a firm) that have wronged.

In the first chapter, I experimentally contrasted the two opposing accounts as to how cheating may affect people's subsequent judgment of the victims in an organizational context. According to the first account, unethical acts may elicit guilt (Cohen et al., 2011; Gneezy et al., 2012; Xu et al., 2012). As a negative moral emotion, guilt may serve as a way to deter people from behaving unethically due to anticipated guilt (Hatch & Kugler, 2019). However, if people already engaged in unethical behavior, guilt may encourage reparative actions and prosocial behaviors in order to "balance the scale" (Cohen et al., 2011; Tracy & Robins, 2007). Prosocial behaviors include compliance with direct requested for help and the willingness to donate time or resources (Shu & Gino, 2012). This account suggests that the shoplifter from the previous example would judge and evaluate the supermarket more favorably to reduce feelings of guilt. Notably, unethical individuals may justify their actions and, therefore, will not regard their actions as unethical. In such cases, cheating will not elicit guilt and subsequent compensatory steps (Hofmann, Wisneski, Brandt, & Skitka, 2014), and therefore will not influence their subsequent judgments.

According to the second account, people sometimes find it difficult to justify their misconduct. They, therefore, distance themselves from their own actions, preferring to harshly judge others so they can view themselves as "ultra-moral" (Barkan, Ayal, Gino, & Ariely, 2012; Shalvi et al., 2015). In other words, people tend to behave consistently with their past moral history and decisions, and so they will act more immorally following an initial immoral behavior (Schwabe, Dose, & Walsh, 2018). Alternatively, they may do so when they want to reassure and, possibly, update their own beliefs about themselves based on their prior actions (Ariely & Norton, 2008). This pattern suggests that people will judge their victim more negatively to preserve a sense of moral self-consistency (Jussim, Yen, & Aiello, 1995). The tendency to judge the victim more negatively is also consistent with the Moral Disengagement theory, which posits that people can selectively disengage moral self-regulation (for instance, by attributing the blame to the victim and their actions) and to make unethical acts more socially or morally acceptable (Bandura, 1999). This opposing account suggests that a shoplifter would judge and evaluate the supermarket more negatively, trying to distance themselves from the object of their transgression.

According to the findings from the first chapter, when people behave unethically, whether their victim is workplace (Studies 1.1-1.3), or governmental authority (Study 1.4), they tend to judge the victim more harshly, presumably to distance themselves and thus justify their transgression. However, it is still unclear if the pattern is relevant only to long-term interactions or also relevant in one-time interactions, such as when using a product or service. Exploring the impact of one-time interactions in the marketing context can be especially insightful since cheating can impact a brand's ability to transform one-time interactions into an affective commitment and behavioral loyalty (Steenhaut & Van Kenhove, 2005).

Moreover, even if people's tendency to distance themselves from the misconduct can be replicated in the marketing realm, the underlying mechanism driving the effect is still unclear. As a self-justification mechanism, distancing is a technique used to decrease ethical dissonance and restore moral self-image (Shalvi et al., 2015). Therefore, the ability to justify unethical behavior can moderate the effect and the ability to judge the victim harshly after behaving unethically.

Distancing as a Self-Justification Mechanism

Research shows that ordinary people who perceive themselves as honest frequently break their own moral code: they lie (almost every day), bend rules, and cut corners for profit (Mazar, Amir, & Ariely, 2008). However, the tension people may feel between two competing motivations: gaining from cheating and maintaining a positive moral self-image, may result in ethical dissonance (Ayal & Gino, 2011; Barkan et al., 2012; Shalvi, Dana, Handgraaf & De Dreu, 2011). In order to avoid anticipated or experienced ethical dissonance, people may use various justifications before or after engaging in unethical behavior (Shalvi et al., 2015). Interestingly, while justifications protect people from psychological tension and restore their moral self-image, they also enable people to behave unethically by distorting their sense of morality (Barkan, Ayal & Ariely, 2015).

Distancing as a self-justification mechanism is a technique used to decrease ethical dissonance and restore moral self-image. Thus, to employ it, people need to have the ability to justify unethical acts. This suggests that the ability to justify can have a moderating role in the impact of unethical behavior on subsequent judgments and behaviors. Meaning, it is possible that when cheating is hard to justify (i.e., cheating is blatant), people may not distance themselves from the act by blaming their victims and instead will attribute their cheating to

themselves or try to atone for their act outside the context of their transgression (Shalvi et al., 2015).

The effect of unethical behavior on distancing may also be moderated by previous behaviors and the possibility of using moral licensing to reconstruct the acts as ambiguous and not immoral (Miller & Effron, 2010). Past research suggests that moral licensing can be achieved using "moral credentials" (Effron, Cameron, & Moni, 2009). Meaning, people who behave morally in one decision may be more likely to construe later moral transgressions as less unethical, having previously established credentials that show they are a moral decision-maker. In this way, people may evaluate that they can act unethically without signaling something morally discrediting about themselves (Miller & Effron, 2010).

An additional factor that can moderate the effect of unethical behavior on subsequent judgments is whether or not the consumer actively sought an advantage or was basically passive. Previous research suggests that consumers tend to believe that it is more unethical to actively benefit from an illegal activity than to benefit from it passively (Vitell & Muncy, 1992). Meaning, consumers tend to think that as long as they do not initiate the action, and instead gain from it passively by saying nothing and letting the seller make the mistake (e.g., getting too much change, receiving an unjustified discount, or given an item for free by mistake) then it is not as wrong or unethical (Muncy & Vitell, 1992).

I designed five experiments to examine if the distancing account can explain how consumers' unethical behavior impacts their judgments in the marketing realm and better understand the underlying mechanism driving the effect and factors that may moderate it. I find that consistent with organizational context, consumers tend to harshly judge those they have cheated (Study 2.1). Providing further support for the distancing account, I find that when cheating became hard to justify, the extent of cheating did not affect judgments (Study 2.2). Additionally, I find support for the moderating role of both moral credentials (Study 2.3) and how consumers tend to increase their profit when passively benefit from cheating (Studies 2.4-2.5). Interestingly, I find no evidence for an effect on guilt or subsequent prosocial behavior (Studies 2.2 and 2.3).

Study 2.1 – How unethical consumers judge a product

In the first study, I examined the common everyday situation where a person can gain a monetary benefit by cheating the seller or vendor of a product they use. To simulate this

situation, I had participants “test” a new product, a mobile app, for an alleged developer and rated it as part of a usability study for the app. Critically, in the cheating condition, participants could use the app unethically to earn more money, while in a control condition, they could not. Controlling participants’ actual earnings, I compared participants’ ratings of the app in the two conditions to determine whether participants’ judgments in the cheating condition would be more positive or more negative following unethical behavior than participants’ judgments in the control condition.

Based on previous studies and results from the first chapter (studies 1.2-1.4), I hypothesized that people who cheat more would more harshly judge the target of their transgression (Mazar et al., 2008; Shalvi et al., 2015). Formally, I predict that:

H2.1 People who earn more money by cheating (cheating condition) will report the lowest satisfaction level with the app compared to people who earn less money by cheating, and compared to people who earn money (low or high amount) without cheating (control condition).

H2.2 The effect of condition (cheating or control) on the general rating of the app will be mediated by the earnings.

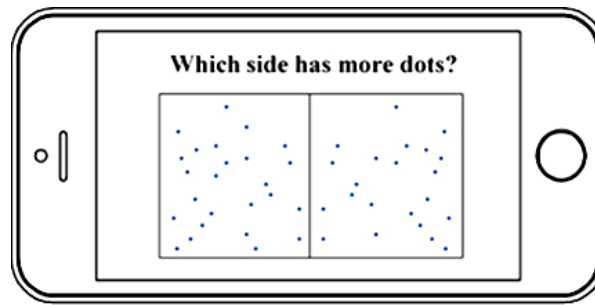
H2.3 There will be a significant negative correlation between earnings and general rating only in the cheating condition.

Method

Participants. I recruited 201 participants (65% male, Mage = 35.0, SD = 7.45) from Amazon Mechanical Turk. Because the use of the app requires accurate vision, I asked participants to confirm that their vision was normal or corrected-to-normal. Participants received a base pay of \$0.05 and a bonus payment of between \$0.20 to \$2, depending on their performance. The study used two incentive conditions (control and cheating), with approximately 100 participants in each condition.

Design and Procedure. Participants were invited to take part in a usability study of a new mobile app using their mobile phones. The app was based on the “dots task,” which previous research has used to measure dishonesty (Gino, Norton, & Ariely, 2010; Hochman, Glöckner, Fiedler, & Ayal, 2016; Kouchaki & Smith, 2014). The app repeatedly displayed, for two seconds, two squares, one containing 25 dots and the other containing 22 dots (Figure 2.1).

Figure 2.1. Image of the mobile app dots task used in Study 1.



Participants were asked to indicate which square contained more dots under two payoff structures (conditions). In the control condition, participants were told they would receive \$0.10 for each correct response and \$0.01 for each incorrect response. By contrast, in the cheating condition, participants were told they would receive \$0.10 if they respond that the right-hand side had more dots and \$0.01 if they respond that the left-hand side had more dots. This payoff structure disregards the actual correct choices and has been found to induce cheating behavior (e.g., Hochman et al., 2016). Nevertheless, in both conditions, participants were explicitly instructed always to choose the side containing more dots. Thus, in the cheating condition, participants knew that a violation of the accuracy instructions would increase their gains unethically. After reading these instructions, participants performed ten practice trials with accuracy feedback. Then they all completed 20 test trials, in a randomized order, under the payoff structure in the condition to which they were randomly allocated. When they completed the task, they were asked to rate how much they liked the app overall (on a 5-star scale), and to also rate the app on nine metrics, including evaluation of its features (graphic design, the difficulty of the task, app's ease of use, responsiveness of the app, and overall performance), and the likelihood of future interaction with the app (play the app again, download the free app when available, willingness to recommend the app to a friend, and purchase an extended version of the app for \$1.99). The nine measures of the app ratings together with the overall liking rating of the app, showed high internal reliability (Cronbach's $\alpha = .87$), so I averaged them to form a composite “general rating” score. Lastly, participants reported their demographics, were thanked for their participation and paid according to their performance in their condition when the study was completed.

Results and Discussion

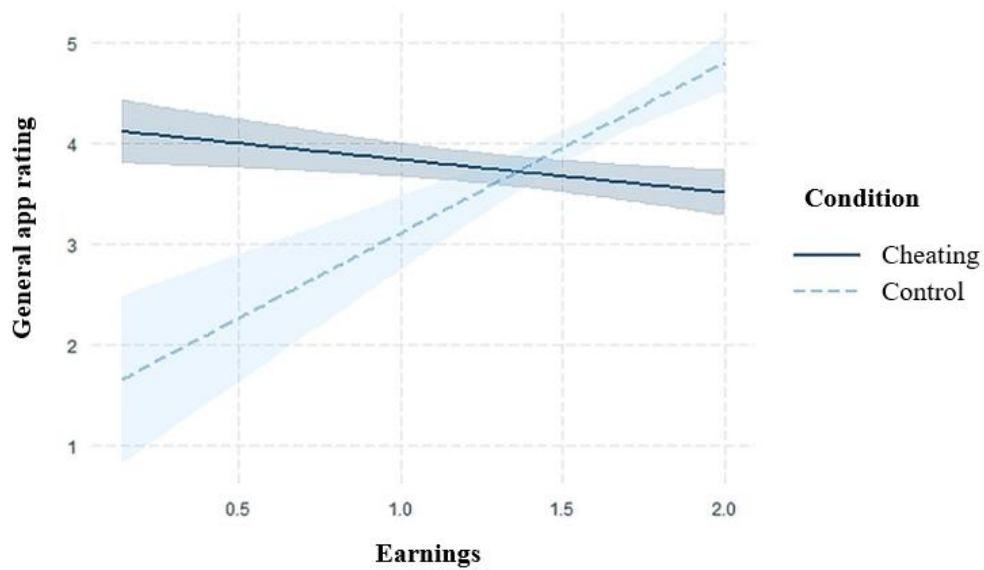
In the study, because the accuracy rates among the control condition participants were higher (78% correct on average) than the cheating rate in the cheating condition (61% claimed

as “right” on average), and because the pay for these two responses was the same (\$.10), earnings were higher in the control condition ($M = \$1.58, SD = .28$) than the cheating condition ($M = \$1.29, SD = .64$), $t(199) = 4.03, p < 0.01$. However, because my interest is in the relationship between the extent of cheating (or earning) and subsequent judgments, this result is inconsequential for my main research question.

I tested the main research question by comparing the relation between earnings and subsequent app judgments in the cheating and control conditions. Higher earnings reflect more cheating in the cheating condition, whereas in the control condition, higher earnings reflect more accurate performance. I used the correlation in the control condition as a baseline for when cheating is not possible. To test hypothesis H2.1, I ran an ANOVA with the amount earned and condition (cheating vs. control group) as independent variables and the general rating score as a dependent variable. Because the earnings level is a continuous variable, I divided the earnings into high and low levels based on the mean of the earnings and the median. There was a significant interaction between the amount earned and condition based on the median of earnings ($M_{\text{earnings}} = 1.62$), $F(1, 197) = 13.73, p < .01$. Participants in the cheating condition with high earnings, reported a low general rating score ($M = 3.28, SD = .88$) compared to participant in the cheating condition with low earnings ($M = 3.68, SD = .78$) and control condition with high and low earnings ($M = 3.99, 3.52; SD = .66, .99$). There was also a significant interaction between the amount earned and the condition based on the mean of earnings ($M_{\text{earnings}} = 1.44$), $F(1,197) = 24.37, p < .01$. However, hypothesis H2.2 was not supported. A mediation analysis using PROCESS method (Model 4 in Hayes, 2017) revealed a non-significant indirect effect of condition (cheating versus control) on the general rating score through earnings (mediator), $b = -.01, 95\% CI [-.07, .08]$.

To further explore the interaction of amount earned and condition on the general rating score, I used the PROCESS method (Model 1 in Hayes, 2017). The analysis revealed a significant moderation effect, $F = 11.53, p < .01$, with a negative coefficient for the interaction term, $b = -1.74, SE = .32, p < .01$. As Figure 2.2 illustrates, and consistent with hypothesis H2.3, there was a significant negative correlation between earnings and general ratings in the cheating condition, $b = -.21, SE = .13, p < .05$, while in the control condition, there was a significant positive correlation, $b = .47, SE = .28, p < .01$. In other words, supporting a distancing account, participants in the cheating condition rated the app less favorably the more they cheated (and earned from it). In contrast, participants in the control condition rated the app more favorably the more they earned from it.

Figure 2.2. The impact of earnings (US\$) on general app rating in cheating and control conditions



Another interesting comparison is to examine judgments in the cheating and control conditions as a function of earnings. As Figure 2.2 shows, participants that cheated significantly (earning more than 1.5 dollars in the cheating condition) rated the app less favorably than those who earned the same amounts of money in the control group, who could not cheat.

Considered jointly, while the mediation effect (hypothesis H2.2) was not supported, it is still evident that the satisfaction level of participants who cheated was significantly lower compared to other participants (hypothesis H2.1). In addition, there was a negative correlation between earnings and satisfaction level (hypothesis H2.3). Meaning, the findings, although not fully supporting the main hypothesis, suggest that cheating was associated with participants subsequently judging the app more negatively. The findings are consistent with the distancing account, in which people judge the object of their transgressions more negatively (Barkan et al., 2012; Shalvi et al., 2015), presumably in order to restore their own moral self-image (Mazar et al., 2008).

Distancing is a technique used to decrease ethical dissonance, and I find evidence for it in Study 2.1, in which the task characteristics presented cheaters with the opportunity to justify the act. However, because distancing is a self-justification mechanism, I speculated that in cases where it is difficult to justify the unethical action (i.e., cheating is obvious), people would not use distancing. The aim of Study 2.2 was to employ a moderation-of-process design to test

the role of distancing by manipulating task difficulty/ease of justification. Moderation-of-process design is being used when it is relatively easy to manipulate the process but difficult to measure (Spencer, Zanna, & Fong, 2005).

Study 2.2 – The moderating role of ease of justification

To explore the moderating role of justification, I used the experimental design of Study 2.1 and modified it to include three levels of discrimination difficulty of the dots task app to manipulate the ease with which participants could justify their cheating (Hochman et al., 2016). Based on previous studies, I hypothesized that justification would moderate the previously found relationship between cheating and rating.

H2.4 In the easy to justify condition (when the task is of hard or medium difficulty), cheating rates will be higher compared to the hard to justify condition (easy task).

H2.5 There will be a significant negative correlation between cheating rates and the general rating of the app only in the easy to justify condition.

H2.6 The effect of condition (ease of justification) on the general rating of the app will be mediated by the cheating rate.

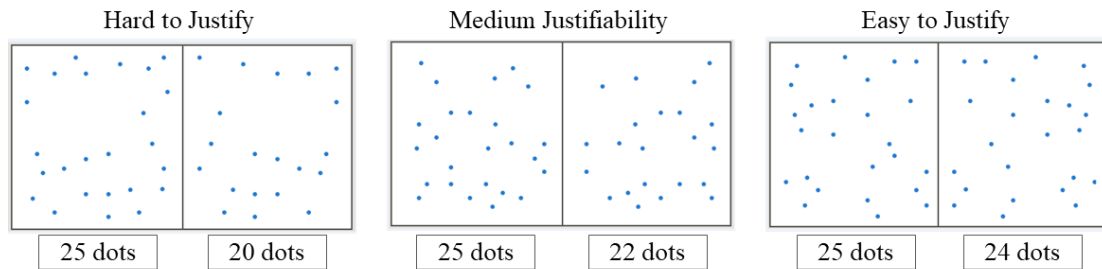
Method

Participants. I recruited 456 participants from Amazon Mechanical Turk (72% female, $M_{\text{age}} = 35$, $SD = 9.7$) who reported having a normal or corrected-to-normal vision. As in Study 2.1, the study was presented as a usability study for a new mobile app. I ensured participants who took part in Study 1 could not take part in this study. Participants earned a base pay of \$0.05 plus a bonus that depended on their performance. The study used a 2 (incentive condition: cheating and control) X 3 (ease of justification: easy, medium, and hard) design, with a minimum of 70 participants in each condition.

Design and Procedure. As in Study 2.1, I used an app that simulates the dots task (Hochman et al., 2016). Departing from Study 2.1, I allocated all participants to the cheating condition and manipulated justifiability (vis-à-vis task difficulty) between-subjects. All participants were tempted to cheat by offering a higher payment for choosing the right-hand side, regardless of whether this side contained more or fewer dots (i.e., \$0.5 for right-hand-side choices and \$0.05 for left-hand-side choices). The square containing more dots always

contained 25 dots. In the hard-to-justify cheating condition, the square with fewer dots contained 20 dots (justification was hard because it was easy to correctly choose the square containing more dots); in the medium-justifiability condition, it contained 22 dots; and in the easy-to-justify condition, it contained 24 dots (Figure 2.3). I adopted these three levels from Hochman et al. (2016), who reported that they indeed exhibit decreasing cheating rates.

Figure 2.3. Image of three levels of justifiability



All participants played the app and then rated it. The survey included a general evaluation of the app (on a 5-star scale) and ratings for six specific features. Participants performed ten practice trials with accuracy feedback and then 25 test trials (in randomized order). Five trials served as fillers and had a higher number of dots on the right-hand side (entered as a covariate in the analyses). After completing the dots task, participants rated the app as in Study 1. The measures showed high internal reliability (Cronbach's $\alpha = .87$), so I averaged them to form a general evaluation score.

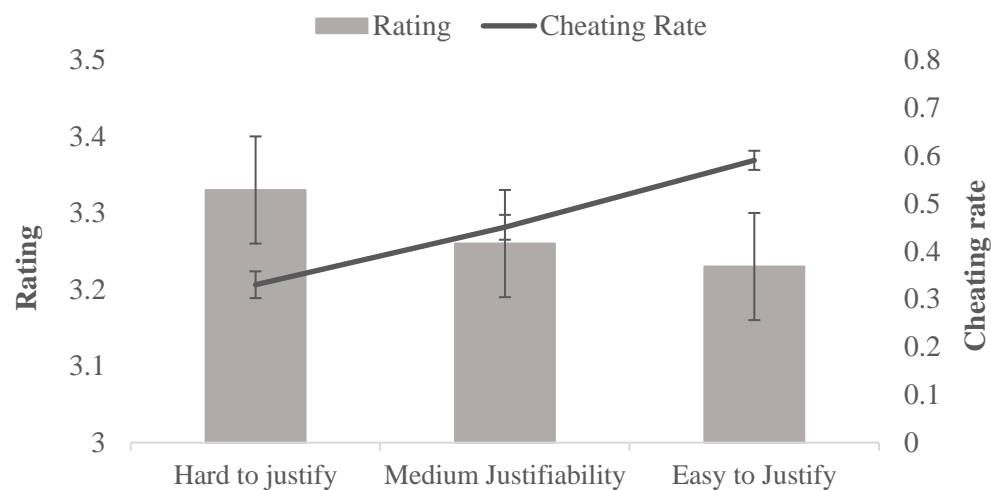
Results and Discussion

Consistent with hypothesis H2.4, I found that the easier it was to justify cheating (the more difficult the task was), the more participants cheated, $F(2, 453) = 24.67, p < .01$. Similar to the findings of Hochman et al. (2016), participants cheated most in the easy-to-justify condition ($M = 0.59, SD = .26$), less in the medium-justifiability condition ($M = 0.45, SD = .33$), and least in the hard-to-justify condition ($M = 0.33, SD = .36$).

Critically, as Figure 2.4 shows, participants in the easy-to-justify condition, who cheated more (and thus earned more money), were least satisfied with the app ($M = 3.23, SD = 0.99$), followed by those in the medium-justifiability condition ($M = 3.26, SD = 1.10$) and those in the hard-to-justify condition ($M = 3.33, SD = 0.98$). A linear regression analysis revealed a negative correlation between cheating level and ratings, $b = -.14, F(1, 454) = 8.25, p < .01$. To control for variance in participants' performance, I used as a covariate the sum of

their correct answers on the ten practice trials and 5 test trials in which there was a higher incentive for choosing the side with more dots in the app. To examine hypothesis H2.5, I Used the PROCESS method (Model 1 in Hayes, 2017), with the condition as a moderator for the effect of cheating rate on rating, I found that the model was significant, $F(4, 451) = 3.32, p < .05$, $F(3, 452) = 3.39, p < .05$ ¹. Importantly, there was a significant and negative correlation between cheating and ratings in both the easy-justifiability condition, $b = -.17, SE = .29, p < .05$, and medium-justifiability condition, $b = -.19, SE = .27, p < .05$, while there was no significant correlation in the hard-to-justify condition, $b = -.10, SE = .22, p = .65$. These findings support hypothesis H2.5.

Figure 2.4. Cheating rate and ratings by ease of justification condition



In order to test hypothesis H2.6 and estimate the direct effect of the condition (ease of justification) on the general rating score, and the indirect effect of the cheating rate as mediator, I performed a mediation analysis using PROCESS method (Model 4 in Hayes, 2017). Results indicated that condition was a significant predictor of cheating rate, $b = .31, SE = .02, 95\% CI [.09, .16], p < .01$, and that cheating rate was a significant predictor of general rating score, $b = -.14, SE = .15, 95\% CI [-.72, -.12], p < .01$. Condition was no longer a significant predictor of general rating score after controlling for the mediator, cheating rate, $b = .01, SE = .06, 95\% CI [-.11, .13], p = .91$, consistent with complete mediation.

The results support my hypotheses. In the easy and medium justifiability conditions, the cheating levels were higher than in the hard-to-justify condition. Critically, only in the easy and medium justifiability conditions were the correlations between cheating level and app

¹ This pattern of results replicated also after removing the covariate.

rating negative. In these conditions, participants who earned more (because they cheated more) rated the app less favorably than those who cheated less. In other words, the target suffered twice: they were cheated and were judged negatively.

In support of the justification account, when justification was relatively easy, participants who cheated more rated the target more negatively. However, when justification was hard because cheating was more obvious, cheaters presumably had more difficulty distancing from the victim and therefore did not rate the target more negatively. Finding that judgments of the target are more negative only when justification is possible supports the distancing-based process as the underlying mechanism driving the effect. Importantly, even in the hard-to-justify condition, in which cheaters could presumably not distance themselves, I found no evidence for the prediction that cheaters would rate the victim more positively because they feel guilty for behaving unethically. Put differently, if participants felt any guilt following their transgression, it did not lead them in this study, or any of the previous studies, to compensate the target of their transgression by rewarding them with more favorable ratings.

Finally, the results of Study 2.2 also help to refute a reversed causality account, whereby a target is evaluated negatively, and therefore people allow themselves to subsequently cheat it. Had this account explained my findings, there should have also been a significant negative correlation between cheating levels and app ratings in the hard-to-justify condition. The fact that there was a significant negative correlation between cheating and ratings only in the medium and easy to justify conditions suggests that cheating drives ratings, and not vice versa.

Study 2.3 – The moderating role of moral credentials

This study aims to explore the moderating role of moral credentials. After behaving morally in one situation, people are more likely to interpret later moral transgressions as less unethical, having previously established credentials that show they are moral decision-makers (Effron et al., 2009). Accordingly, I predict that:

H2.7 People with a high accuracy rate would tend to cheat more only in the hard to justify condition.

H2.8 There will be no significant correlation between the cheating rate and general rating due to moral credential.

H2.9 The effect of condition (ease of justification) on the general rating of the app will be mediated by the cheating rate.

To explore the moderating role of moral credentials, I used the experimental design of Studies 2.1 and 2.2 and modified it to include two parts – the first part included an incentive for accuracy, and the second part included an incentive for cheating. While the first app provided the possibility to have high accuracy rates (medium degree of difficulty), the second app varied to an easy-to-justify condition and high-to-justify condition to measure the effect in both cases.

In addition, in this study, I examine whether people tend to engage in prosocial behavior following unethical behavior. According to previous studies, misbehavior may evoke guilt feelings (Cohen et al., 2011) and encourage prosocial behaviors in order to repent for the sins and "balance the scale" (Tracy & Robins, 2007). Prosocial behaviors include compliance with requests for help and a higher willingness to donate time or resources (Shu & Gino, 2012). Importantly, findings in previous studies in this research did not reveal any prosocial behavior after cheating. Therefore, I predict that:

H2.10 There will be no correlation between the cheating rate and prosocial behavior.

Method

Participants. I recruited 189 participants from Amazon Mechanical Turk (51% female, $M_{\text{age}} = 32$, $SD = 8.9$) who reported having a normal or corrected-to-normal vision. As in Study 2.1, the study was presented as a usability study for a new mobile app. Participants earned a base pay of \$0.05 plus a bonus that depended on their performance. The study used two conditions (easy and hard to justify), with approximately 100 participants in each condition.

Design and Procedure. As in the last two studies, I used an app that simulates the dots task (Hochman et al., 2016). However, in this study participants were asked to test two versions of the app (within-subjects design) and review 25 screens of each version. The first version was similar to all participants and had no incentive for cheating (incentive per accurate answers). After completing the task, participants were asked to rate the app (on a 5-star scale) and were assigned to a second version of the app. In the second app, participants were tempted to cheat by offering a higher payment for choosing the right-hand side, regardless of whether this side contained more or fewer dots (i.e., \$0.5 for right-hand-side choices and \$0.05 for left-hand-side choices). Participants were randomly assigned to an easy-to-justify condition or

hard-to-justify condition. In all the versions, the square with more dots contained 25 dots. In the control version (no incentive for cheating), the square with fewer dots contained 22 dots (medium difficulty level). In the Cheating versions, I used the same manipulation as in Study 2.2 (based on Hochman et al., 2016). Namely, in the hard-to-justify cheating condition, the square with fewer dots contained 20 dots (justification was hard because it was easy to correctly choose the square containing more dots); and in the easy-to-justify condition, it contained 24 dots. As in Study 2, five screens in the cheating versions served as fillers and had a higher number of dots on the right-hand side (entered as a covariate in my analyses). To make the first and second versions of the app visually different, the squares in the control version had two background colors (unlike the white background color in the cheating condition) – the left square was blue, and the right square yellow.

After participants completed the second version of the app, they were requested to rate this version as well (on a 5-star scale). Finally, to explore the willingness to engage in prosocial behavior towards the target of their transgression, participants were asked if they are willing to donate their time to review some additional screens (with no extra payment) to detect bugs in the system. Participants were asked how much time (0 seconds – 180 seconds) they are willing to donate to check those additional screens.

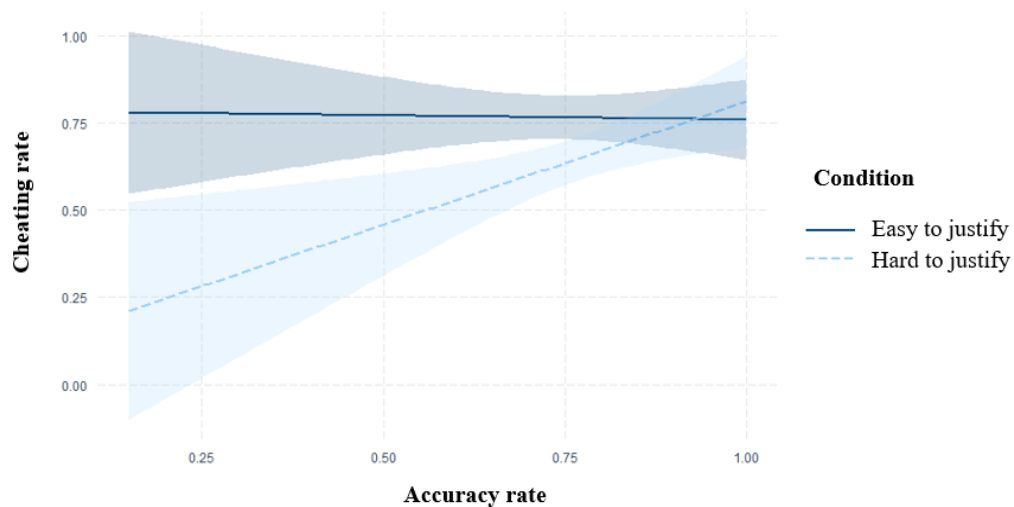
Results and Discussion

Cheating and Accuracy rates. Both cheating and accuracy rates were calculated based on the 20 screens that varied between the versions and represented either cheating or accuracy of participants' answers. The average accuracy rate in the control condition was 0.76 ($SD = .14$), with a median of 0.79. The average cheating rate in the cheating condition was 0.70 ($SD = .37$), with a median of 1.00 (50% of all participants cheated in all the screens to maximize their profits).

As in Study 2.2, I found that similar to the findings of Hochman et al. (2016), participants cheated more in the easy-to-justify condition ($M = 0.76$, $SD = .30$), and less in the hard-to-justify condition ($M = 0.65$, $SD = .42$), $t(187) = 2.20$, $p < 0.05$. Interestingly, using the PROCESS method (Model 1 in Hayes, 2017), with the condition as a moderator for the effect of accuracy rate on cheating rate, I found that the model was significant, $F = 3.42$, $p < .05$. with a negative coefficient for the interaction term, $b = -.71$, $SE = .37$, $p < .05$. As Figure 2.5 illustrates and consistent with hypothesis H2.7, there was a significant positive correlation between accuracy rate and cheating rate in the hard-to-justify condition, $b = .69$, $SE = .29$, $p <$

.05, while in the easy-to-justify condition, there was a non-significant negative correlation, $b = -.02$, $SE = .22$, $p = .91$. Meaning, the cheating rates in the easy-to-justify were higher and not affected by the previous performance, most probably since justification was available also without moral credentials. However, in the hard-to-justify condition, participants probably used previous performance to justify cheating since it could be portrayed as an ambiguous and not immoral act (Miller & Effron, 2010).

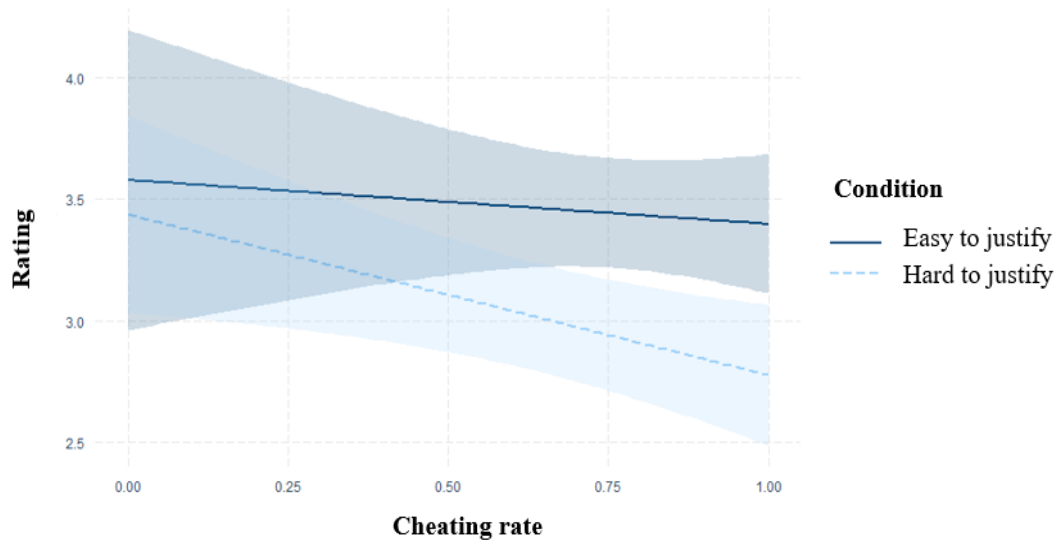
Figure 2.5. The impact of performance (accurate rate) in the control condition on the cheating rate in easy-to justify and hard-to-justify conditions.



Rating. The average rating of the app in the control condition ($M = 3.66$, $SD = .98$), was significantly higher than the cheating condition ($M = 3.22$, $SD = 1.33$), $t(189) = 4.39$, $p < .01$. In addition, the average rating of the app in the easy-to-justify condition ($M = 3.79$, $SD = 1.00$) was significantly higher than the hard-to-justify condition ($M = 3.52$, $SD = .95$), $t(187) = 1.86$, $p < .05$. Using the PROCESS method (Model 1 in Hayes, 2017), with the condition as a moderator for the effect of cheating rate on rating, I found that the model was significant, $F = 3.19$, $p < .05$. As Figure 2.6 illustrates, there was a significant negative correlation between the cheating level and rating in the hard-to-justify condition, $b = -.21$, $SE = .31$, $p < .05$, while in the easy-to-justify condition, there was a non-significant negative correlation, $b = -.04$, $SE = .46$, $p = .69$. Meaning, hypothesis H2.8 was not supported. A possible explanation for these findings is that participants cheated more due to moral credentials in the hard-to-justify condition. However, it might be that moral credential was still not enough to justify the cheating and so participants judged the app less favorably in order to distance themselves. Yet, in the

easy-to-justify condition, participants were able to fully justify the cheating, and therefore their ratings were not affected by the cheating level.

Figure 2.6. The impact of cheating rates on rating in easy-to cheat and hard to cheat conditions



In order to test hypothesis H2.9 and estimate the direct effect of the condition (ease of justification) on the general rating score, and the indirect effect of the cheating rate as a mediator, I performed a mediation analysis using PROCESS method (Model 4 in Hayes, 2017). Results indicated that condition was a significant predictor of cheating rate, $b = .32$, $SE = .05$, 95% $CI [.01, .22]$, $p < .05$, and that cheating rate was a significant predictor of general rating score, $b = -.14$, $SE = .26$, 95% $CI [-1.02, -.01]$, $p < .05$. The indirect effect of the condition on the general rating score through cheating was significant, $b = -.04$, 95% $CI [-.12, .00]$.

Prosocial behavior. The average time participants were willing to donate in order to review more screens was 56.3 seconds ($SD = 58.7$), with no significant difference between the easy-to-justify condition ($M = 49.8$, $SD = 54.8$) and the hard-to-justify condition ($M = 62.5$, $SD = 61.9$), $t(187) = 1.48$, $p = 0.14$. In order to test hypothesis H2.10, I used a linear regression analysis. The analysis revealed no significant correlation between the cheating level and willingness to donate additional time, $b = -.06$, $F(1, 187) = 0.72$, $p = .39$. Using a PROCESS method (Model 1 in Hayes, 2017), with satisfaction level as a covariate, I did not find a significant moderation effect, $F = 2.08$, $p = .09$. In addition, both conditions had no significant correlation between the cheating rate and willingness to donate additional time to review the app. The findings support hypothesis H2.10.

The results support the moderating role of moral credentials. While previous studies (e.g., Hochman et al., 2016) demonstrated a lower cheating rate among participants in the hard-to-justify conditions, the findings of this study suggest that previous performance affected the cheating rate. Meaning, participants with high accuracy rates in the control condition tended to cheat more in the subsequent task. This is consistent with the notion that people who previously established credentials that show they behaved morally are more likely to engage later on in unethical behavior (Miller & Effron, 2010), even if cheating is obvious, as in our study. Importantly, contrary to the findings in Study 2.2, participants who cheated more in the hard-to-justify condition rated the app less favorably than those who cheated less. Meaning, while moral credentials enabled participants to justify the transgression and see it as less unethical, the dishonesty was not fully justified and led to distancing from the target of the transgression.

In addition, there was no correlation between previous performance and cheating rate in the easy-to-justify conditions, presumably since participants did not need to establish moral credentials when justification was already available. However, contrary to findings in Study 2.2, participants' subsequent judgment of the app was not affected by their cheating rates. The results suggest that moral credentials enabled participants to construe their later misbehavior as less immoral (Efron et al., 2009) and fully justify it. Lastly, although previous studies suggest that misbehavior may encourage prosocial behavior (Tracy & Robins, 2007), I found no indication of prosocial intentions in both conditions. Meaning, even when participants cheated more, their willingness to engage in prosocial behavior did not change. These findings reinforce the results from Study 2.2, in which there was no evidence for guilt and subsequent reparative actions.

Study 2.4 - The moderating role of actively benefiting from cheating: pretest

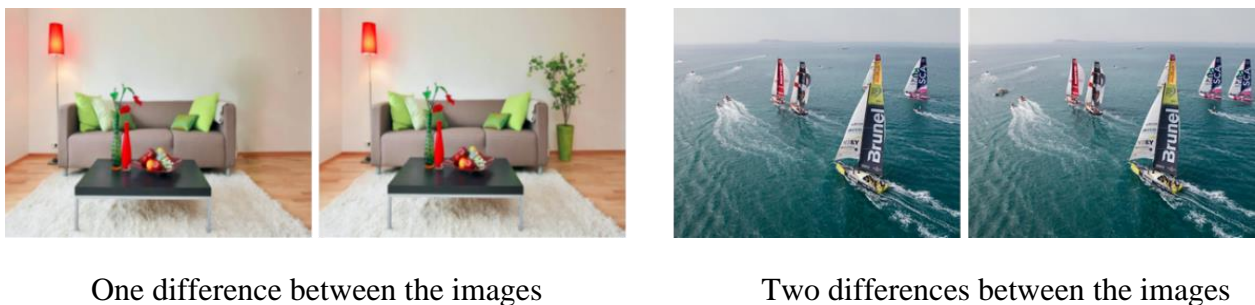
In the next set of experiments, I used a different task to check the moderating effect of actively benefiting from unethical acts. The *spot-the-difference task* (Gai & Puntoni, 2021) involves finding exactly three differences between the two images that are simultaneously presented for a few seconds. Participants received a fixed reward (0.10\$) each time they reported that they found precisely three differences (self-report). Unbeknown to participants, there are three types of difference trials: three-, two-, and one- difference trials. Therefore, claims of having found three differences reflect cheating when there are less than three differences between the two images. The pretest aimed to determine the amount of time needed to evaluate the images.

Method

Participants. I recruited 109 participants from Prolific (58% female, $M_{\text{age}} = 34$, $SD = 10.1$). Because participants were required to identify differences based on small details and color changes, participants were asked to confirm that their vision was normal or corrected-to-normal and that they do not suffer from color blindness. Participants earned a total of £0.40 for the study. The study was similar to all participants, and therefore had a minimum of 100 participants.

Design and Procedure. I used the *spot-the-difference task* (Gai & Puntoni, 2021). In the pretest, I presented participants with ten pairs of images with either one, two, or three differences between them (Figure 2.7). Participants were informed about the number of differences between each pair of images and were asked to identify the differences as fast as possible and to click on the 'Next' button when they do.

Figure 2.7. Example of the Spot-the-difference task images



Results and Discussion

The pretest included five pairs of images with three differences, three pairs of images with two differences, and three pairs of images with one difference. The average response time was significantly different between the screens according to the number of differences in each pair, so that it was higher as the number of differences increases (Table 2.1).

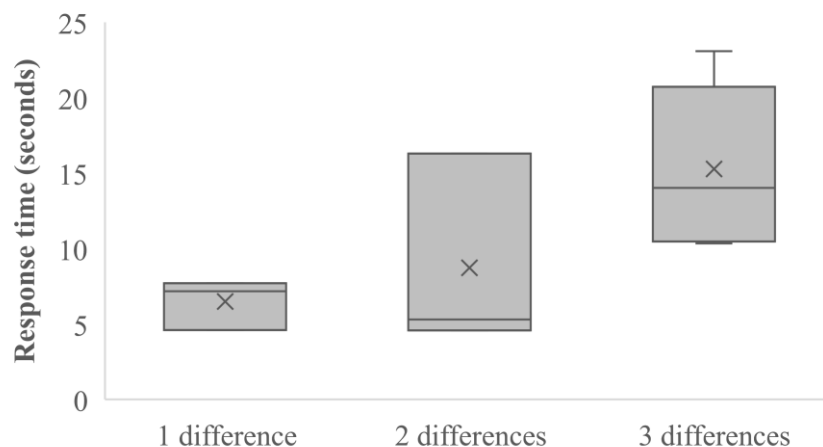
Table 2.1. Response time (Mean) according to the number of differences between images

	M	SD	95% Credible Interval	
			Lower Bound	Upper Bound
One difference	10.13	12.87	7.67	12.60
Two differences	11.99	9.39	10.19	13.79

Three differences	20.35	16.76	17.13	23.56
-------------------	-------	-------	-------	-------

The median of the response time varied between the pairs and was higher as the differences between the images increased (Figure 2.8). Based on the response time from this pretest, I set a response time of 15 seconds for the subsequent study to provide enough time to review the screens and detect the differences, but at the same time to limit the response time to the minimum necessary to encourage cheating (Shalvi et al., 2012). In addition, one pair of images was eliminated from the pool of questions since the response time was too high ($M = 32.5$ seconds, $SD = 27.3$, $Mdn = 23$), and participants seemed to have difficulties in finding the differences.

Figure 2.8. Response time (Median) according to the number of differences between images



While participants identified the differences between the images, the data in the pretest suggest that the time in which participants find the differences may vary significantly between participants. Meaning, in the following study, when a time limit is set, participants may falsely think they have cheated when indicating they found three differences. Similarly, the difficulty in finding three differences may cause participants to think that cheating is too obvious and therefore reduce the cheating level in the cheating condition. Therefore, it is possible that the results in the following study will include a low level of cheating and will consequently make it difficult to draw clear conclusions.

Study 2.5 – The moderating role of actively benefiting from cheating

In this study, I examined the moderating role of actively versus passively benefiting at the expense of others. According to Muncy & Vitell (1992), consumers' beliefs on questionable

acts are influenced by whether they actively sought advantage at others' expense or were basically passive. Specifically, consumers tended to believe that it was more unethical to actively benefit from dishonesty than to passively benefit (Vitell, 2003). To simulate this situation, I used the spot-the-difference task, with a crossed 2 (cheating, control) X 2 (passive, active) design. I manipulated the ability to gain more money actively or passively by determining the default answer as a "yes" (passive) or "no" (active).

H2.11 People in the "passively cheating" condition will have a higher cheating rate compared to people in the "actively cheating" condition.

H2.12 There will be a positive correlation between cheating and rating in the passive cheating condition and a negative correlation between cheating and rating in the active cheating condition.

H2.13 The effect of condition (cheating or control) on the rating will be mediated by cheating rates.

Method

Participants. I recruited 393 participants from Prolific Academic (51% female, $M_{\text{age}} = 31$, $SD = 10.8$). As in Study 2.4, I asked participants to confirm that their vision was normal or corrected-to-normal and that they do not suffer from color blindness. Participants earned a base pay of £0.10 plus a bonus depended on performance (up to an additional £1.00 bonus payment). The study used a 2 (cheating and control) X 2 (passive vs. active benefit from dishonesty) design, with approximately 100 participants in each condition.

Design and Procedure. As in Study 2.4, I used the *spot-the-difference task* (Gai & Puntoni, 2021). The task involves finding three differences between the two images that are simultaneously presented for 15 seconds. Within this time frame, participants were asked to report whether they found exactly three differences between the two images ("Yes" button to indicate they found three differences and the "No" button to indicate they did not find three differences). Participants reviewed ten different pairs of images and earned a fixed reward (£0.10) each time they reported that the images had three differences. The task had two types of difference trials: three and two differences trials. Therefore, each claim of having found three differences reflects cheating when there were only two differences between the images.

The study was a crossed 2 (cheating/control) x 2 (passive/active) between-subjects design. The first two conditions were the ability to cheat: in the control condition, participants received pairs with three differences, and so there was no incentive to cheat to gain more money; In the cheating condition, most of the pairs (8 out of 10) were with only two differences and therefore presented an opportunity to cheat to increase gains. Each of the conditions included two possibilities – "yes" as the default answer (meaning, participants were informed that if they would not respond, the system would automatically code their response as a "yes" and they will receive the reward). In the second condition, the "No" was the default answer (meaning, if participants did not answer after 15 seconds, the system coded the answer as "no" and participants did not receive their reward). These two conditions represented, in the cheating condition, two possibilities to cheat in order to increase gains – passively benefiting from cheating ("yes" as a default answer) and actively benefiting from cheating ("no" as a default answer).

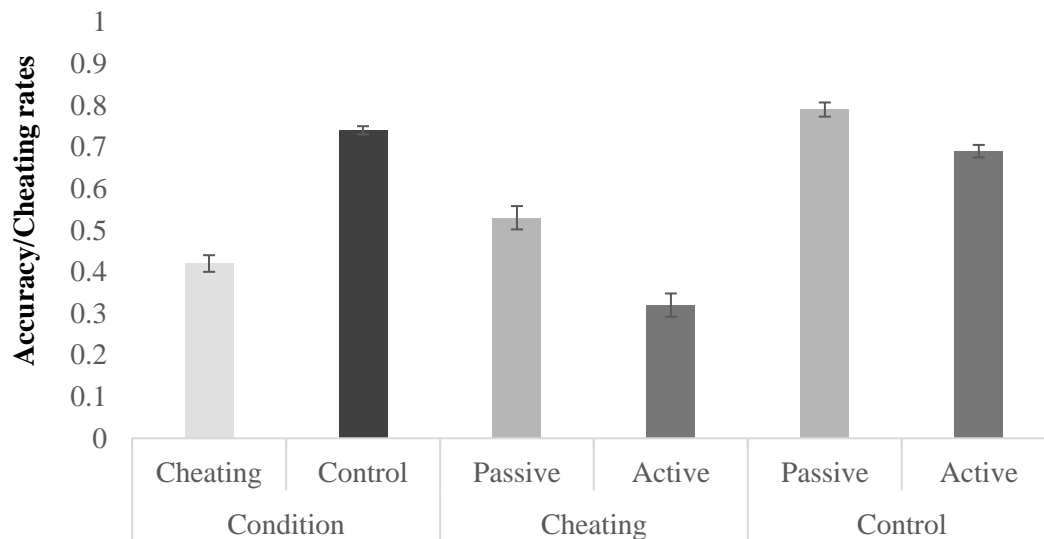
Participants first read the task instructions and completed one practice trial with three differences between the images. After completing the task in one of the four versions (cheating/control X passive/active), participants were requested to rate the overall liking of the app (on a 5-star scale) and four features of it: the graphic design, the app's ease of use, the willingness to play with the app again, and recommend the app to a friend. The four measures of the app ratings together with the overall liking rating of the app, showed high internal reliability (Cronbach's $\alpha = .81$), so I averaged them to form a composite "general rating" score. Lastly, participants were asked whether they think that people might feel justified to over-report their actual number of times they could find three differences on a scale of 1-5 (1 – not at all, 2 – unlikely, 3 – maybe, 4 – probably yes, 5 – definitely yes).

Results and Discussion

Cheating/Accuracy rate. In both conditions (cheating and control), two pairs of images were identical and had three differences. The two pairs served as a covariate to control for variance in participants' performance. In the control condition, the additional eight pairs also had three differences and therefore measured the accuracy rate. In the cheating condition, the additional eight pairs had only two differences between the two images, and therefore any response of "yes" for those was considered cheating. As Figure 2.9 illustrates, the accuracy rates (control condition) and cheating rates (cheating condition) were significantly different, $M = 0.74, 0.42$; $SD = .17, .30$; $t(394) = 12.95, p < .01$. in addition, consistent with hypothesis

H2.11, the cheating rates were significantly different between the passive cheating (yes as a default) and the active cheating (no as default) in the cheating condition, $M = 0.53, 0.32$; $SD = .28, .28$; $t(193) = 5.28, p < .01$, and the control condition, $M = 0.79, 0.69$; $SD = .17, .16$; $t(199) = 4.26, p < .01$.

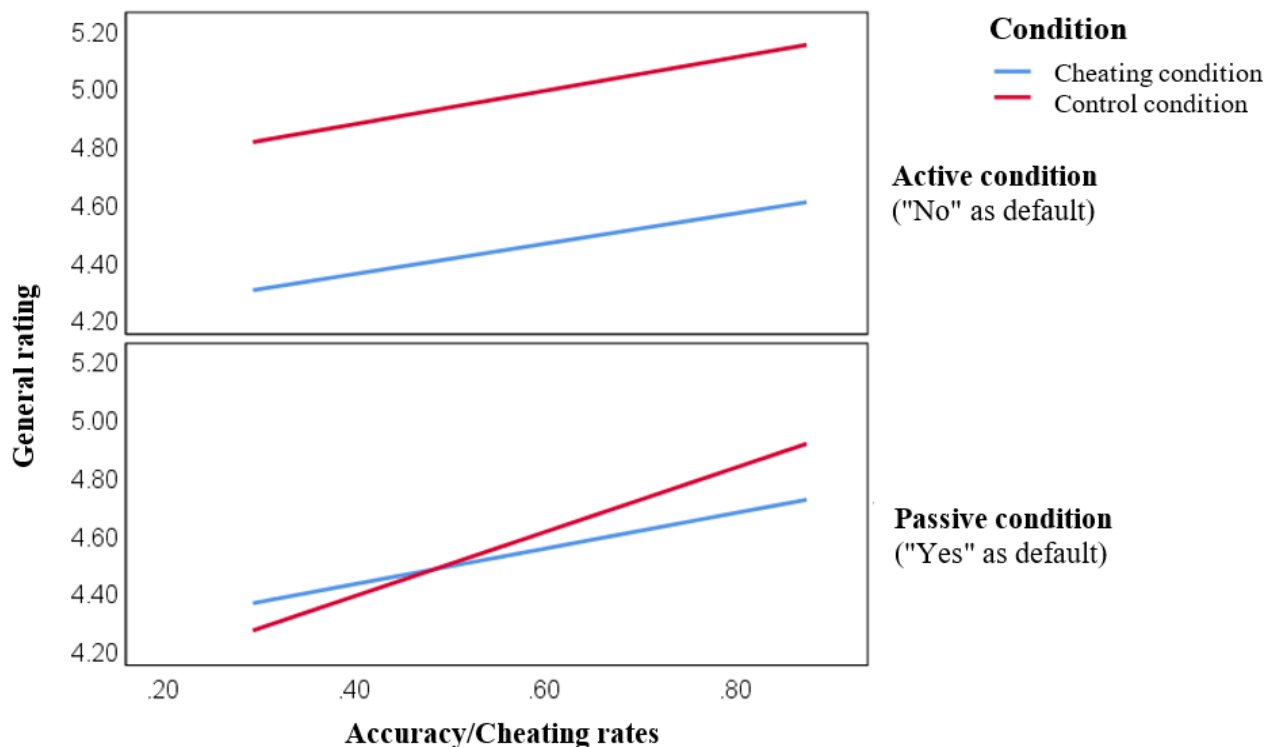
Figure 2.9. Accuracy/cheating rates according to condition (cheating and condition) and default answer (passive and active).



General rating. The rating was significantly higher in the control condition ($M = 4.92, SD = .81$) compared to the cheating condition ($M = 4.41, SD = .95$), $t(394) = 5.77, p < 0.01$. However, in the cheating condition, there was no significant difference in the general rating between the passive condition ($M = 4.51, SD = .95$) compared to the active condition ($M = 4.30, SD = .95$), $t(394) = 1.50, p = 0.13$. Using the PROCESS method (Model 3 in Hayes, 2017), with the two identical pairs as a covariate, revealed that the condition significantly moderates the effect of accuracy/cheating rate on general rating, $F(8,387) = 7.03, p < .01$. As Figure 2.10 illustrates, consistent with hypothesis H2.12, there was a significant positive correlation between cheating rate (cheating condition) and general ratings in the passive condition, $b = .20, SE = .04, p = .05$. In addition, there was a significant positive correlation between accuracy rate (control condition) and general rating in the passive condition, $b = .24, SE = .06, p < .05$. However, there was no significant correlation between cheating rate and general rating in the active condition, $b = .17, SE = .04, p = .10$, and between accuracy rate (control condition) and general rating in the active condition, $b = .14, SE = .06, p = .30$. Meaning, when participants could gain benefits in a passive way, either by taking advantage at the expense of others or not, they were more satisfied as the success rates increased. However,

when participants had to gain more money actively, they did not rate the app differently based on their success rates.

Figure 2.10. The impact of accuracy/cheating rate on general app rating in cheating and control conditions



In order to test hypothesis H2.13 and to estimate the direct effect of the condition (ease of justification) on the general rating score, and the indirect effect of the accuracy/cheating rate as a mediator, I performed a mediation analysis using PROCESS method (Model 4 in Hayes, 2017). Results indicated that condition (cheating or control) was a significant predictor of accuracy/cheating rate, $b = 1.12$, $SE = .19$, 95% $CI [2.24, 2.98]$, $p < .01$, and that accuracy/cheating rate was a significant predictor of general rating score, $b = .20$, $SE = .02$, 95% $CI [.03, .12]$, $p < .05$. The indirect effect of the condition on the general rating score through cheating was significant, $b = .22$, 95% $CI [.08, .37]$.

Justification to over-report. The average score of justification to over-report was significantly higher in the cheating condition ($M = 4.26$, $SD = .75$) compared to the control condition ($M = 4.11$, $SD = .85$), $t(394) = 1.81$, $p < 0.05$. While there was no significant difference in the score of justification to over-report between the passive and active conditions when I used as a covariate the cheating rate, there was a negative correlation between the general rating and justification to over-report score both cheating conditions: in the passively cheating condition, $b = -.20$, $F(1,95) = 3.76$, $p < .05$, and in the actively cheating condition, b

= -.19, $F(1,96) = 3.55$, $p < .05$. However, in the control conditions, the correlation was not significant. Interestingly, there was no significant correlation between cheating or accuracy rates (whether passive or active) and justification to over-report in the app.

The results suggest that when cheating, participants are more willing to take passive advantage of others than participants who need to gain more benefits actively (Muncy & Vitell, 1992). Also, findings support the moderating role of active versus passive benefiting from cheating, since participants, based on their cheating level, rated the app differently in the two conditions. When passively cheating, participants rated the app more favorably as cheating rates were higher. A possible explanation of the positive correlation is that people tend to believe that it is less unethically to passively benefit from dishonesty (Vitell, 2003). Therefore, participants who did not initiate dishonesty were more satisfied as their (justified) gains were higher. Another possible explanation is that even though participants were passively benefiting at the expense of others, they felt guilty for doing it and therefore compensate the victim by providing a more favorable rating (Shu & Gino, 2012). Based on the negative correlation between rating and the beliefs regarding whether people can justify over-reporting, the guilt account may better explain the effect. Namely, when participants thought people would not feel justified to over-report the actual cases they found, they provided a higher ranking. On the contrary, when they felt that most probably people will justify over-reporting, they felt less obliged to rate the app favorably. This may suggest that even if the company made the mistake, participants compensate it by providing higher ranks.

Contrary to hypothesis H2.12, the correlation between the cheating rate and the rating of the app was not significant in the actively cheated condition. A possible explanation is that the need to actively change the default answer in order to get more money made the cheating more obvious and, therefore, harder to justify. Like in Study 2.2, in this case, participants found it difficult to blame the victim. Another possible explanation for the results is that many participants did not cheat (the Median of cheating was two screens out of eight, with only 20% of participants cheating in more than four screens), which may have created a floor effect.

General Discussion

The research enables a better understanding of the important and understudied phenomenon: how consumers judge those they have cheated. In five experiments, I show that when consumers behave unethically, they tend to judge the victim harsher, compared to when they cannot act unethically, and that is presumably done in order to distance themselves, and thus justify their transgression. The first study explored the impact of unethical behavior on

subsequent judgments of a new app. The results were not consistent and did not support the mediation role of cheating (H2.2). However, the findings suggest that people who cheat rate the target of their transgression significantly lower (H2.1), and more negatively the more they cheat (hypothesis H2.3). The result suggest that the negative consequence of unethical acts is likely to be quite pervasive because it happens not only in long-term relationships, such as with an employer but also in one-time interactions, such as when using a particular product or app.

Also, demonstrating an important boundary condition, I find that the unethical act does not influence judgments when cheating is “blunter” and hard to justify (Study 2.2). Findings that judgments are less favorable only when the misbehavior can be justified (hypotheses H2.5-H2.6) support the distancing-based process as the mechanism that drives the effect. Meaning, since distancing is a self-justification mechanism, people should be able to use it only when justification is possible and accordingly judge the victim of their transgression in order to restore their moral self-image (Shalvi et al., 2015). As in Studies 1.3 and 1.4, this study’s results refute a reversed causality account whereby a prior negative judgment allows people to cheat more. If that was the case, one should have found a significant negative correlation between cheating rate and subsequent rating also in the hard to justify condition.

In addition, I find support for the moderating role of moral credentials (Study 2.3; hypothesis H2.7), in which people tend to cheat more after behaving morally (Effron et al., 2009). Findings suggest a higher cheating rate, although cheating was hard to justify after performing well in the control condition. Meaning, contrary to previous studies that indicate that people tend to cheat less when it is harder to justify the misbehaviors (Hochman et al., 2016) and do not negatively judge the target of their transgression after cheating (Study 2.2), moral credentials provided the ability to justify the cheating and use a distancing account as a post justification mechanism. However, while previous studies emphasized the role of actively benefiting from cheating (Muncy & Vitell, 1992), I find no evidence for the moderating role of it on the impact of unethical behavior on subsequent judgments of the app (Study 2.5; hypothesis H2.12). Findings suggest that while cheating rates are higher when passively benefiting on others' expense, there was no significant difference in the subsequent rating whether the benefit was acquired actively or passively. Importantly, findings suggest that the app's overall subsequent rating was significantly lower when given the opportunity to cheat.

The research adds to the extant literature by considering not only the direct costs of unethical behavior (e.g., lost revenues) but also its less direct effects on customers’ judgments

and word of mouth. As a result, it may impact long-term marketing strategies and brand perception. Importantly, it can affect brands' efforts to transform one-time interactions into effective communication (Steenhaut & Van Kenhove, 2005) and a consumer-brand relationship to encourage positive emotions and behavioral loyalty (Ghani, 2016; Khamitov, Wang, & Thomson, 2019). In addition, it may highlight the importance of developing business and marketing strategies that consider the downside effect of unethical behavior. These strategies may highlight the importance of making it harder to justify unethical acts or adding reminders that emphasize the active role consumer have in cheating actions. Also, organizations can use creative methods throughout the customer journey to minimize the possible damage to the firm's perception.

In addition, the research informs both the theory and practice of online reviews (Babić, Sotgiu, De Valck, & Bijmolt, 2016; Chen, & Kirmani, 2015; Chevalier, & Mayzlin, 2006) by exposing a novel factor that can systematically sway ratings. Given that unethical consumption appears to be common, with up to 64% of consumers engaging in unethical consumption at some point in their lives (Farmer & Dawson, 2017), it is likely that many online reviews originate from unethical consumption. The findings suggest that these reviews may be biased because unethical consumers may distance themselves from the acts to restore their moral self-image. However, the extent of this phenomenon is unclear because unethical consumers may choose not to write reviews. Future studies might examine the direct and indirect impacts of unethical consumption on the review market.

Lastly, the research suggests implications for policies regarding consumer protection and fair trade. Since unethical behavior is relevant in different settings (Ayal & Gino, 2011) and may impact judgments, behaviors, and word of mouth, it is important to inform people about the possible bias in judgments while interacting with governmental authorities, business organizations, brands, and workplaces. Also, findings suggest that organizations and brands may face significant consequences due to unethical behavior, which policymakers may need to consider going forward. However, to consider policies or interventions that aim to protect customers and firms, further studies are required to understand the phenomenon and the boundary conditions better.

Limitations and directions for future research

The research examines the impact of unethical behavior on subsequent judgments and behaviors in a marketing context. Although the results in study 2.1 were not conclusive in terms

of mediation of the cheating rate on the effect of the condition on the general rating score, the findings in other studies (2.2-2.5) supported the distancing account. While a possible explanation for the lack of mediation effect in study 2.1 can be attributed to the small sample ($N = 201$), additional studies should examine the distancing response in other tasks used in the behavioral ethics field (e.g., matrix task, roll a die).

In addition, while the results suggest a distancing-based process as the underlying mechanism driving the tendency of people to judge their victims harshly, a better understanding of the mechanics of the distancing process is needed – whether people tend to distance themselves to relieve ethical dissonance, a moral self-regulation disengagement to reframe the act as less immoral or attribution of blame to the victim to avoid self-reproach feelings. In addition, while the ability to withstand threats to the self-concept through moral disengagement may lead to post-violation justification (Shalvi et al., 2015), it is also possible that using moral disengagement strategies and strategic forgetting of moral rules to make unethical acts personally acceptable (Shu, Gino, & Bazerman, 2011) can actually reduce ethical dissonance and, consequently, the need to justify the action by judging the victim harshly.

Future research may aim to reevaluate the moderating role of actively benefiting from unethical acts (Muncy & Vitell, 1992). While the findings of Study 2.5 did not support my hypothesis, it is possible that the task being used (spot-the-difference task; Gai & Puntoni, 2021) or the low cheating rates affected the results. Additional studies can try to better understand the effect and the role of active versus passive cheating. Another finding in the research that can be explored in the future is the willingness to engage in prosocial behavior after behaving immorally (Tracy & Robins, 2007). While I find no evidence of it in Study 2.3, a further examination might explore the willingness to engage in prosocial behavior in cases where other justifications are not available (such as distancing in my studies). Furthermore, I did not manipulate other potential moderating factors such as the size of the potential victim's company (Wirtz and McColl-Kennedy 2010) and the victim's identifiability (Gino et al. 2010). Future studies could examine the effect of transgression toward small, identified companies (e.g., local grocery shop, café) versus large companies (e.g., supermarket chain) on subsequent judgments.

Finally, while I tested the impact of unethical behavior on subsequent ratings, I did not address the possible effect on written reviews. Many online reviews are posted on social media and internet forums (where star ratings are not present) and influence people's behavior and

firm perception (Zhong & Schweidel, 2020). In addition, verbal reviews contain rich information about the experience and consumers' sentiments and may differ from a mere rating (Büschken & Allenby, 2016). However, it is unclear whether subsequent written reviews of consumers are consistent with their rating or not. Chapter 3 will address this research question and examine the effect of dishonesty on the sentiments that consumers express, shaping people's behavior and preferences.

Conclusion

In conclusion, this research explores the consequence of unethical behavior in a marketing context, that is, how consumers judge those they have cheated. Behaving unethically may encourage consumers to distance themselves and harshly judge the company, providing lower ratings and potentially causing further harm to those they have already cheated. In addition, the research suggests that the ability to justify unethical behavior serves as the underlying mechanism of the phenomenon. This research offers important implications for marketers, managers, and policymakers seeking to influence organizations and firms' perceptions and maintain brand equity.

References

- Ariely, D., & Norton, M. I. (2008). How actions create—not just reveal—preferences. *Trends in cognitive sciences*, 12(1), 13-16.
- Association of Convenience Stores (2019). *The crime report 2019*. Retrieved from https://www.acs.org.uk/sites/default/files/acs_crimereport2019_online_version.pdf
- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. Mikulincer M, Shaver PR, eds. *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (American Psychological Association, Washington, DC) 149-166.
- Ayal, S., Hochman, G., & Ariely, D. (2016). Dishonest behavior, from theory to practice. *Frontiers in Psychology*, 7, 1521.
- Babić, Rosario. A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757-773.
- Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, 6(DEC), 157-161.
- Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6), 953-975.
- Business Software Alliance (2018). *2018 BSA Global Software Piracy Study*. Retrieved from <https://gss.bsa.org/>.
- Chen, Y. J., & Kirmani, A. (2015). Posting strategically: The consumer as an online media planner. *Journal of Consumer Psychology*, 25(4), 609-621.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.

- Cohen, TR., Wolf, ST., Panter, AT., & Insko, CA. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, 100(5), 947-966.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of experimental social psychology*, 45(3), 590-593.
- Farmer, J. F., & Dawson, J. (2017). American College Students' Shoplifting Experience: A Comparison of Retrospective Self-Reports to Micro-Level Criminological Theory. *International Journal of CriminFal Justice Sciences*, 12(1), 1.
- Fombelle, P. W., Voorhees, C. M., Jenkins, M. R., Sidaoui, K., Benoit, S., Gruber, T., ... & Abosag, I. (2020). Customer deviance: A framework, prevention strategies, and opportunities for future research. *Journal of Business Research*, 116, 387-400.
- Fullerton, R. A., & Punj, G. (2004). Repercussions of promoting an ideology of consumption: consumer misbehavior. *Journal of Business Research*, 57(11), 1239-1249.
- Gai, P. J., & Puntoni, S. (2021). Language and consumer dishonesty: a Self-Diagnosticity theory. *Journal of Consumer Research*.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1-44.
- Ghani, N. H. B. A. (2016). Consumer brand relationships. *International Review of Management and Marketing*, 6(4).
- Gino, F., & Ariely, D. (2016). Dishonesty explained. What leads moral people act immorally. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 322–344). New York, NY: Guilford Press.
- Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological science*, 21(5), 712-720.
- Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless+ harmless= blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes*, 111(2), 93-101.
- Gneezy, U., Imas, A., & Madarász, K. (2012). Conscience accounting: Emotional dynamics and social behaviour. *Theoretical economics paper series* (STICERD, London).

- Hatch, D., & Kugler, T. (2019). From Unethical to Prosocial Behavior: The Role of Guilt, Positive Affect, and System Processing. In *Academy of Management Proceedings* (Vol. 2019, No. 1, p. 11732). Briarcliff Manor, NY 10510: Academy of Management.
- Hayes, AF. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (Guilford Press, New York).
- Hochman, G., Glöckner, A., Fiedler, S., & Ayal, S. (2016). “I can see it in your eyes”: Biased processing and increased arousal in dishonest responses. *Journal of Behavioral Decision Making*, 29(2/3), 322-335.
- Hofmann, W., Wisneski, DC., Brandt, MJ., & Skitka, LJ. (2014). Morality in everyday life. *Science*, 345(6202), 1340-1343.
- Jussim, L., Yen, H., & Aiello, JR. (1995). Self-consistency, self-enhancement, and accuracy in reactions to feedback. *Journal of Experimental Social Psychology*, 31(4), 322-356.
- Khamitov, M., Wang, X., & Thomson, M. (2019). How well do consumer-brand relationships drive customer brand loyalty? Generalizations from a meta-analysis of brand relationship elasticities. *Journal of Consumer Research*, 46(3), 435-459.
- Kouchaki, M., & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological science*, 25(1), 95-102.
- Levav, J., & McGraw, AP. (2009). Emotional accounting: How feelings about money influence consumer choice. *Journal of Marketing Research*, 46(1), 66-80.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6), 633-644.
- Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. In *Advances in experimental social psychology* (Vol. 43, pp. 115-155). Academic Press.
- Muncy, J. A., & Vitell, S. J. (1992). Consumer ethics: An investigation of the ethical beliefs of the final consumer. *Journal of business Research*, 24(4), 297-311.
- National Retail Federation (2020). *Consumer Returns in the Retail Industry 2020*. Retrieved from <https://nrf.com/research/customer-returns-retail-industry>

- NRF - National Retail Federation (2019). *2019 Organized retail crime report*. Retrieved from <https://cdn.nrf.com/sites/default/files/2019-12/NRF%20ORC%20Survey%202019.pdf>
- Potdar, B., Guthrie, J., Gnoth, J., & Garry, T. (2018). The role of psychological ownership in shoplifting prevention: an exploratory study. *Journal of Retailing and Consumer Services*, 44, 253-265.
- PwC – PriceWaterhouseCooper (2018). *PwC's Global economic crime and fraud survey 2018*. Retrieved from <https://www.pwc.com/gx/en/forensics/global-economic-crime-and-fraud-survey-2018.pdf>.
- Ruedy, NE., Moore, C., Gino, F., & Schweitzer, ME. (2013). The cheater's high: The unexpected affective benefits of unethical behavior. *Journal of Personality and Social Psychology*, 105(4), 531-548.
- Schwabe, M., Dose, D. B., & Walsh, G. (2018). Every Saint has a Past, and Every Sinner has a Future: Influences of Regulatory Focus on Consumers' Moral Self-Regulation. *Journal of Consumer Psychology*, 28(2), 234-252.
- Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181-190.
- Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, 23(10), 1264–1270.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125-130.
- Shu, LL., & Gino, F. (2012). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*, 102(6), 1164-1177.
- Shu, LL., Gino, F., & Bazerman, MH. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3), 330-349.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of personality and social psychology*, 89(6), 845.

- Steenhaut, S., & Van Kenhove, P. (2005). Relationship commitment and ethical consumer behavior in a retail setting: The case of receiving too much change at the checkout. *Journal of Business Ethics*, 56(4), 335-353.
- Taylor, E. (2016). Supermarket self-checkouts and retail theft: The curious case of the SWIPERS. *Criminology & Criminal Justice*, 16(5), 552-567.
- Tracy, J.L., & Robins, R.W. (2007). The self in self-conscious emotions: A cognitive appraisal approach. Tracy J.L., Robins R.W., Tangney J.P., eds. *The Self-Conscious Emotions: Theory and Research* (Guilford Press, New York), 3-20.
- Vitell, S. J. (2003). Consumer ethics research: Review, synthesis and suggestions for the future. *Journal of business ethics*, 43(1-2), 33-47.
- Vitell, S. J., & Muncy, J. (1992). Consumer ethics: An empirical investigation of factors influencing ethical judgments of the final consumer. *Journal of Business Ethics*, 11(8), 585-597.
- Vitell, S. J., & Muncy, J. (2005). The Muncy–Vitell consumer ethics scale: A modification and application. *Journal of Business Ethics*, 62(3), 267-275.
- Wirtz, J., & McColl-Kennedy, J.R. (2010). Opportunistic customer claiming during service recovery. *Journal of the Academy of Marketing Science*, 38(5), 654-675.
- Xu, H., Bègue, L., & Bushman, B.J. (2012). Too fatigued to care: Ego depletion, guilt, and prosocial behavior. *Journal of Experimental Social Psychology*, 48(5), 1183-1186.
- Yaniv, G. (2009). Shoplifting, monitoring and price determination. *The Journal of Socio-Economics*, 38(4), 608-610.
- Zhong, N., & Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science*.

Chapter 3

The Impact of Reviews by Unethical Individuals on Other People's Decisions

Abstract

Although ample research has examined the impact of online written reviews on other people's decisions and behaviors, it is still unclear whether reviews written after unethical behavior differ in their influence on readers. Meaning, while previous chapters in this research examined the impact of unethical behavior on subsequent judgments, further examination is needed to understand the possible influence of those judgments on other people's behaviors and their willingness to engage with the victim. In two studies, I test the differences in written review sentiments between participants who cheated or behaved ethically and whether the different reviews influence other people's willingness to use the product. In the first study, I use written reviews collected in previous studies (Studies 2.1 and 2.2) and find support for the distancing account with more negative sentiments after cheating, but only when there is a possibility to justify the cheating (Study 3.1). In addition, I find that the willingness to use a product after reading a review depends on whether the review was written by people who cheated or not (Study 3.2). These findings suggest important implications in both marketing and organizational contexts because cheating may lead to possible bias in online reviews and thus impact the perception and image of brands, organizations, and public service. In addition, policies dealing with consumer protection and fair trade may need to examine ways to approach possible bias in online reviews while considering a service or product.

Introduction

People often share opinions, news, and information with others. The development of online social media platforms (e.g., Facebook, Instagram, Tumblr) and e-commerce websites (e.g., Amazon, eBay, Walmart) provided people with the possibility to share information about goods, services, and brands (Babić, Sotgiu, De Valck, & Bijmolt, 2016; Blackshaw & Nazzaro 2006; Chen, & Kirmani, 2015). People share their experience with a new restaurant, gossip about their workplace, discuss political issues, or complain about bad government services. As so, online reviews have become a popular source of information and influence others' opinions and decision-making (Berger, 2014; Chen & Yuan, 2020; Kudeshia & Kumar, 2017; Moore & Lafreniere, 2020).

However, while previous research explored online reviews' role in shaping others' opinions about a product or service (Kudeshia & Kumar, 2017), research has not examined the impact of unethical behavior on the online written review's nature and sentiments. More specifically, although unethical behavior is prevalent in different settings, such as organizations, governmental authorities, and the marketplace (Ayal & Gino, 2011), it is still unclear if unethical behavior impacts the subsequent judgments of the victim and eventually other people's decisions and behaviors. Studies show that people tend to read the text of multiple online written reviews (Chevalier and Mayzlin, 2006; Murphy, 2019) and that their perception of organizations and government is based not only upon their personal experience but also on word of mouth (Houston, Aitalieva, Morelock, & Shults, 2016). Meaning, unethical behavior may have far-reaching implications in terms of the perception and reputation of organizations. It may impact the public trust and the ability of the government to implement public policies (Gordon, 2000) and even jeopardize the efforts of firms to recruit skilled employees.

Online Reviews

Studies have shown that word-of-mouth (WOM) communication is more persuasive than traditional marketing actions because it is perceived as independent and more trustworthy (Ivanova & Scholz, 2017; Nielsen, 2015; Trusov, Bucklin, & Pauwels, 2009). With the rise of the Internet and online communication, WOM took the form of online reviews (also referred to as electronic word-of-mouth, eWOM; Ismagilova, Slade, Rana, & Dwivedi, 2019) and became accessible worldwide with more reviews and opinions compared to close personal contacts (Reimer & Benkenstein, 2016). People are increasingly relying on online reviews for

their search of information and view them not only as guidance for purchasing everyday products, such as books or movies, but also in more essential decisions such as medical and financial products (Hu, Bose, Koh, & Liu, 2012; Kostyra, Reiner, Natter, & Klapper, 2016; Packard & Berger, 2017).

Although online reviews refer to both ratings (usually based on a 5-point scale) and text reviews, there is a distinct difference between the two types (Chen & Yuan, 2020; Jia, 2018). While useful, the star rating is a straightforward indicator of performance. However, written reviews provide room for people to express sentiments, describe their experience and the extent to which they are satisfied with the product or service (Jia, 2018). Further, on a typical review platform, star ratings tend to converge as more reviews are posted, making it less meaningful for people that find it challenging to rely solely on the average star rating (Zhang, Zhao, Cheung, & Lee, 2014).

In addition, star rating and written review may be inconsistent in terms of positive and negative opinions. It is possible for a star rating to be extremely high or low while the written review be mixed, with both negative and positive text, or even contrary to what is expressed in the star rating. Lack of match between star ratings and text may indicate that the sender is not an expert who can accurately describe the product (Chen & Yuan, 2020; Moore & Lafreniere, 2020). Hence, inconsistencies between a reviewer's rating and written review will likely reduce the review's perceived helpfulness by raising concerns about the reviewer's ability and even their willingness to tell the truth (Schlosser, 2011).

Notably, while consumer written reviews can be considered an extension of the ratings since they allow people to express opinions on different aspects of a product or service (Büschken & Allenby, 2016), they can sometimes be found separately without a star rating. The growth of digital usage has created new platforms such as blogs, social media, and internet forums, where people can share their opinions towards firms, brands, or events with a significant effect on firms or brands image, customer's preferences, and purchase intention (Chuang, 2020; Kauffmann, Peral, Gil, Ferrández, Sellers, & Mora, 2019). For example, recent studies show that nearly fifty percent of the shoppers reported reading text reviews before making a purchase decision (Tata, Prashar, & Gupta, 2020).

Despite evidence that written reviews are not necessarily consistent with star ratings and that people read the text of online written reviews rather than rely solely upon average star rating (Chevalier and Mayzlin, 2006), much of the research until recently focused on the star

rating and tended to ignore the written content due to the costs and efforts in measuring it (Babić et al. 2016; Forman, Ghose, & Wiesenfeld, 2008; Netzer, Feldman, Goldenberg, & Fresko, 2012; Sen & Lerman, 2007). However, recent research shows that written reviews contain sentiments and information not evident in the star rating and are important while seeking a better evaluation of products and making decisions (Kauffmann et al., 2019; Ziegele & Weber, 2015).

Written Reviews and Sentiments

According to recent studies, people usually read the text of multiple reviews, typically more than ten items, before trusting a firm or brand or making a purchase decision (Murphy, 2019; Varga & Albuquerque, 2019). People consider linguistics cues and style, including the length of the written review, narrative, valence, sentiments, and word use while reading the reviews and determine how trustworthy they are. For example, people who wrote negative reviews with dispreferred markers (such as, “I’ll be honest,” or “I don’t want to be mean, but...”) are perceived to be more credible and persuasive compared to people who wrote negative posts without these markers (Chen & Yuan, 2020; Moore & Lafreniere, 2020).

Many past studies have emphasized the effect of online reviews' valence or sentiments on people's attitude toward the firm and purchasing behavior (Berger et al., 2020; Reimer & Benkenstein, 2016). A review's valence refers to whether the review's direction is positive, neutral, or negative (Lee & Youn, 2009). Positive sentiments affect emotional trust, attitudes toward a brand or firm, and purchasing intentions (Hsu, Yu, & Chang, 2017; Kudeshia & Kumar, 2017). In contrast, negative text reviews can send negative signals and decrease brand equity or purchase likelihood (Bambauer-Sachse & Mangold, 2011; Varga & Albuquerque, 2019). Yet, negativity bias in which people are more influenced by negative information, is limited in the online context, since readers suspect that negative reviews are generated by competitors (Babić et al., 2016), or are the outcome of dissimilar tastes or even bad luck (Chen & Yuan, 2020).

Importantly, written reviews and customer-generated content indicate people attitudes and preferences, influence the image of organizations, and impact retailer's revenue (Kumar, Shankar, & Aljohani, 2020; Moore & Lafreniere, 2020; Packard & Berger, 2017; Rahman & Khamparia, 2016; Zhong & Schweidel, 2020). The importance of customer-generated content expands beyond business organizations and concerns government agencies and policymakers (Bertot, Jaeger, & Hansen, 2012). Citizens use media networks to express themselves about

the different services that their governments are delivering, their opinions on policies, and issues related to their day-to-day situations. Policymakers can no longer ignore these new sources of information since the published information in these networks can influence the governments' perception in the general public's eyes (Bertot et al., 2012; Driss, Mellouli, & Trabelsi, 2019; Houston et al., 2016). In addition, Studies have shown that people's perceptions may influence public trust in the government, and eventually, the ability to implement public policies and achieve cooperative compliance of the citizens (Gordon, 2000; Im, Cho, Porumbescu, & Park, 2014).

However, the main challenge in written reviews is that while star rating data are structured, written reviews consist of natural language and therefore are unstructured and difficult to understand (Jia, 2018). While user-generated content contains immense knowledge about people's decision-making, data about communication between firms and their consumers, and among consumers themselves, there is a need to convert this raw material into valuable insights (Berger et al., 2020; Humphreys & Wang, 2018). Within this context, sentiment analysis techniques are a useful way to examine opinionated text. Sentiment analysis is the process of automatically extracting feelings from the text (Pang & Lee, 2008) and identify the mood or opinion of subjective aspects within the text (Bhadane, Dalal, & Doshi, 2015). Specifically, Sentiment Analysis in reviews explores the text to determine the overall opinion or feeling about a product or a firm (Kauffmann et al., 2019) and the sentiment strength of it ranging from very negative to very positive emotions (Villarroel Ordenes, Ludwig, De Ruyter, Grewal, & Wetzels, 2017).

Due to the increasing importance of user-generated content and written reviews, there is a growing interest in extracting sentiments and insights among marketers, managers, and policymakers (Chen & Yuan, 2020; Chevalier & Mayzlin, 2006). Studies indicate that 72% of companies invest in content analysis (The CMO survey, 2020; Zhong & Schweidel, 2020), and more than half of C-level executives describe word of mouth and text analysis as a key business priority (Simonson & Rosen 2014). However, it is still unclear how unethical behavior impact written reviews and the sentiments expressed in them. Research in eWOM did not address the possible downstream effects of unethical behavior and its potential role in shaping people's judgments, behaviors, and word of mouth.

The Impact of Unethical Behavior on written Reviews

Unethical behavior is a common phenomenon that can be found in different settings in daily life - shoplifting, tax evasion, and over-claiming insurance are some examples of acts that ordinary people commit regularly (e.g., Ayal & Gino, 2011). Although the accumulative economic impact of unethical behaviors is significant (Gerlach, Teodorescu, & Hertwig, 2019), only a few studies addressed the subsequent consequences of these acts and how they may impact the judgment of those that were cheated. Previous studies, presented in Chapters 1 and 2, suggested that people who behave unethically tend to distance themselves from the act, preferring to criticize others so they can present themselves as "ultra-moral" and restore their moral self-image. However, it is not clear if this tendency is also relevant when people who cheated write a review about the victim of their transgression and whether their written review influences other people's behavior and willingness to engage with the victim in a similar way to those generated after behaving ethically.

Written reviews have unique characters that may distinguish them from star ratings; they can include a mix of both negative and positive sentiments, sometimes not consistent with the star rating (Chen & Yuan, 2020), they provide room to express opinions and sentiments (Jia, 2018), add information above the straightforward averaged star rating (Zhang, Zhao, Cheung, & Lee, 2014), and can be found separately from star rating on social media platforms and internet forums (Chuang, 2020). Given the different features of written reviews and the importance it has on people's emotional trust, attitudes toward a firm or a product, and purchasing intentions (Adamopoulos, Ghose, & Todri, 2018; Babić et al., 2016; Hsu et al., 2017), it is essential to have a better understanding concerning the impact of unethical behavior on reviews' sentiments and possible other cues that may influence readers (Chen & Yuan, 2020). Such cues can include, for example, using a higher total level of words and less first-person singular, which are typical to people lying or behaving unethically (Tausczik & Pennebaker, 2010).

The research findings may have far-reaching implications not only in the marketing realm but also for organizations, workplaces, and governmental authorities (Berger, 2014; Driss et al., 2019). People share their experience and opinions about their workplace, the political situation, the service they received while renewing their driving license, and their recently paid taxes. Thus, if the distancing account will be relevant to written reviews as well, it might have an impact on the reputation and image of the organization or public service, it may decrease trust in governmental authorities and even jeopardize the possibility to attract and maintain high-quality employees (Gordon, 2000; Im et al., 2014).

Due to their nature, the sentiments expressed in them and other linguistic cues, written reviews may reveal an opposite outcome of unethical behavior on subsequent judgment compared to the star ratings. Although not found in my previous studies (Chapters 1 and 2), research in the behavioral ethics field suggest that unethical behaviors may elicit guilt (Cohen, Wolf, Panter, & Insko, 2011; Gneezy, Imas, & Madarász, 2012) and encourage reparative actions and prosocial behaviors. Prosocial behavior may serve to "balance the scale" and resolve the tension people feel after misbehaving (Tracy & Robin, 2007). The guilt account suggests that people who misbehaved may include positive sentiments in their written reviews, highlight products' positive features, and even recommend it. Importantly, since written reviews often appear after star rating, the distancing and harsh judgment of the victim after unethical behavior may evoke guilt. As a result, and to repent for their negative star rating, people may provide more positive written reviews implying the importance of the order in which reviews are posted online.

However, according to the distancing account, people sometimes find it difficult to deny their misconduct and therefore tend to harshly judge others in an attempt to distance themselves from their own actions and view themselves as "ultra-moral" (Barkan, Ayal, Gino, & Ariely, 2012; Shalvi, Gino, Barkan, & Ayal, 2015). In other words, according to the distancing account, people who misbehaved may use mostly negative sentiments in their written reviews and highlight negative aspects of the experience or product in order to distance themselves from the target of their transgression and restore their moral self-image.

In two studies, I test these two predictions. In the first study, I analyzed the written reviews collected in two previous studies (Studies 2.1 and 2.2). Text analysis of the reviews reveals that, similar to star ratings, participants tend to show a higher level of negative sentiments after cheating, but only when it is possible to justify their acts. In the second study, I explore whether reviews written by consumers who cheated or not, affect people's willingness to use the product. Previous studies indicate that while negative or positive sentiments in the written reviews are important, other cues, such as the length of the review and the words used, may influence the level at which the review is perceived as trustworthy and persuasive (Chen & Yuan, 2020; Tausczik & Pennebaker, 2010). My findings suggest that people who read reviews written by participants who cheated were less inclined to use the product compared to reviews written by participants who did not cheat.

Study 3.1 – How unethical consumers review a product?

In the first study, I examined whether the sentiments in written reviews differ between participants who cheated compared to participants who behaved ethically and whether the ability to justify the cheating moderates the effect and the sentiments in the written reviews. Based on previous studies' findings (Chapters 1-2), I hypothesize that people who cheat will tend to harshly judge the target of their transgression by using more negative sentiments in an attempt to distance themselves and restore their moral self-image, but only when self-justification is possible.

H3.1 There will be a significant positive correlation between earnings and negative sentiments in the cheating condition.

H3.2 The effect of condition (cheating or control) on negative sentiments will be mediated by earnings.

H3.3 There will be a positive correlation between cheating rates and negative sentiments except when justification is hard.

H3.4 The effect of condition (the ability to justify the behavior) on negative sentiments will be mediated by cheating rates.

H3.5 There will be a negative correlation between cheating rates and positive sentiments in the easy to justify condition.

H3.6 The effect of condition (the ability to justify the behavior) on positive sentiments will be mediated by cheating rates.

In this study, I used written reviews that were collected as part of two dots task studies in chapter 2 (Studies 2.1 and 2.2). In the dots task studies, I asked participants to write a review about the app and provide as many details as possible about the app they tested and their experience. Responses from the same participants for both star ratings and written reviews enabled a simulation of real-life incidents, in which people are requested to provide both rating and written reviews.

Database

The database included 657 written reviews from Studies 2.1 and 2.2 (Chapter 2). Both studies included a request from participants to add their written reviews about the app. The

request was made to analyze the sentiments, language being used, and other cues besides star ratings. While the data from Study 2.1 was used to understand the differences in sentiments between participants who cheated compared to participants that did not cheat, the data from Study 2.2 was used to address the possibility that the ability to justify the unethical behavior may influence the sentiments expressed in the written reviews.

Written reviews collected in Study 2.1. In this study, 201 participants (65% male, $M_{age} = 35.0$, $SD = 7.45$) were recruited from Amazon Mechanical Turk and confirmed that their vision was normal or corrected-to-normal. Participants were requested to take part in a usability study of a new mobile app using their mobile phones. The app was based on the dots task (Gino, Norton, & Ariely, 2010; Hochman, Glöckner, Fiedler, & Ayal, 2016; Kouchaki & Smith, 2014) in which participants were requested to choose, between two squares, the one containing more dots.

The app had two payoff structures (conditions). In the control condition, participants were paid \$0.10 for each correct response and \$0.01 for each incorrect response. By contrast, in the cheating condition, participants received a higher incentive when choosing the right-hand side, although it had fewer dots. This payoff structure disregards the actual correct choices and has been found to induce cheating behavior (e.g., Hochman et al., 2016). Notably, in both conditions, participants were instructed always to choose the side with more dots. Meaning, participants in the cheating condition knew that a violation of the accuracy instructions would increase their gains unethically. When participants completed the task, they were asked to rate how much they liked the app overall (on a 5-star scale) and rate nine app metrics (e.g., graphic design, app's ease of use, and responsiveness of the app). In addition to the star rating, participants were requested to write a review of the app and be as detailed as possible. Participants were obliged to write at least five characters and were not limited in the review length.

Written reviews collected in Study 2.2. In this study, 456 participants (72% female, $M_{age} = 35.0$, $SD = 9.7$) were recruited from Amazon Mechanical Turk. Participants confirmed that their vision was normal or corrected-to-normal. As in Study 2.1, the study was presented as a usability study for a new mobile app that simulates the dots task (Hochman et al., 2016). However, in this study, all participants were allocated to the cheating condition with differences in the task difficulty (manipulation of the ability to justify the cheating) between-subjects. All participants were tempted to cheat by offering a higher payment for choosing the right-hand

side, regardless of whether this side contained more or fewer dots. While the square containing more dots always contained 25 dots, the square with fewer dots varied in the number of dots according to the ease of justification. In the hard-to-justify cheating condition, the square with fewer dots contained 20 dots (justification was hard because it was easy to correctly choose the square containing more dots); in the medium-justifiability condition, it contained 22 dots; and in the easy-to-justify condition, it contained 24 dots. The three levels were adopted from Hochman et al. (2016), who reported that they indeed exhibit decreasing cheating rates. After completing the dots task, participants rated the app (on a 5-star scale) and were asked to write a review on the app with as many details as possible. As in Study 2.1, participants were obliged to write at least five characters and were not limited in the written review length.

Text Analysis Tools

To analyze the written reviews, I used the LIWC2015 (Linguistic Inquiry and Word Count) software that was found to be accurate in identifying emotions in language use (Tausczik & Pennebaker, 2010). According to previous studies, the software rating of positive emotion words (e.g., love, nice, sweet) and negative emotion words (e.g., hurt, ugly, nasty) correspond with human ratings of the writing excerpts (Kahn, Tobin, Massey, & Anderson, 2007). The software dictionary identifies the most common words in the human language and defines each in one or several word categories (for example, negative emotion can be simultaneously categorized in three different categories - sadness, negative emotion, and overall affect). The score is calculated by comparing the words being used to a list of dictionary words that define each category scale. For example, if LIWC analyzed a single speech with 2,000 words, it compares them to the built-in dictionary and might find that there are 150 pronouns and 84 positive emotion words used. Therefore, it would convert these numbers to percentages, 7.5% pronouns and 4.2% positive emotion words (Pennebaker, Booth, & Francis, 2007).

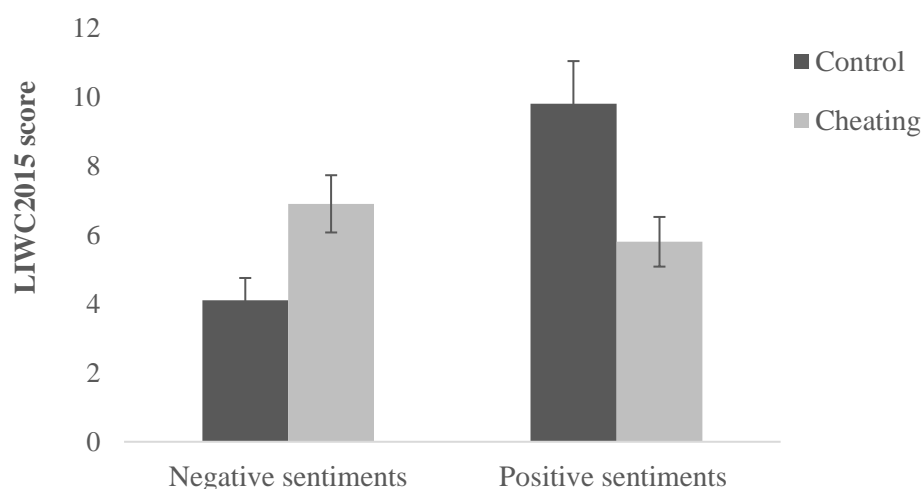
To better understand the text's overall sentiment tone and validate the sentiments' direction in the written reviews (negative or positive), I also used the MonkeyLearn.com software. MonkeyLearn.com is a platform that uses machine learning to get relevant data from text and was found to provide accurate sentiment classification due to its high-quality attributes result and a low result of mean square error (MSE) (Basmimi, Abd Halim, & Saadon, 2020). MonkeyLearn.com focuses not only on polar sentiments (positive, negative) but also on feelings and emotions (angry, happy, sad), urgency (urgent, not urgent), and even intentions (interested v. not interested).

Results and Discussion

Positive sentiments in the written reviews included terms such as *nice*, *fun*, *exciting*, *like*, *engaging*, *interesting*, *enjoyable*, *entertaining*, and even *addicting*. In comparison, the negative sentiments included words such as *boring*, *pointless*, *hard*, *confusing*, *tedious*, *frustrating*, *annoying*, *stressful*, and *anxious*. There were very few neutral reviews, such as "It was fine" or "I wasn't quite sure what counted as a "difference" in a few pictures." In addition, part of the reviews included both sentiments to describe participants' experience and impression from the app (for example, "The app was fun but slightly boring, and I honestly didn't understand the point of it." Or "An interesting way to pass the time but would get redundant very fast. Not the type of brain teaser game I would choose to play").

Cheating and control conditions. There was a significant difference between the control and cheating conditions in terms of positive and negative sentiments according to LIWC software (Figure 3.1): There were significantly more positive sentiments in the control condition compared to the cheating condition ($M = 9.77, 5.83$; $SD = 12.28, 7.32$; $t(199) = 2.78$, $p < .01$). However, this was reversed in the negative sentiments, with significantly more negative terms in the cheating condition compared to the control condition ($M = 6.90, 4.10$; $SD = 8.47, 6.49$), $t(199) = 2.62$, $p < .01$. In other words, supporting a distancing account, participants in the cheating condition used more negative terms and expressed harsher judgment toward the product compared to participants in the control condition. These findings were consistent with the MonkeyLearn analysis, indicating that the control group's general sentiment tone is 72.2% positive, while the cheating group's sentiment tone is 70.8% negative.

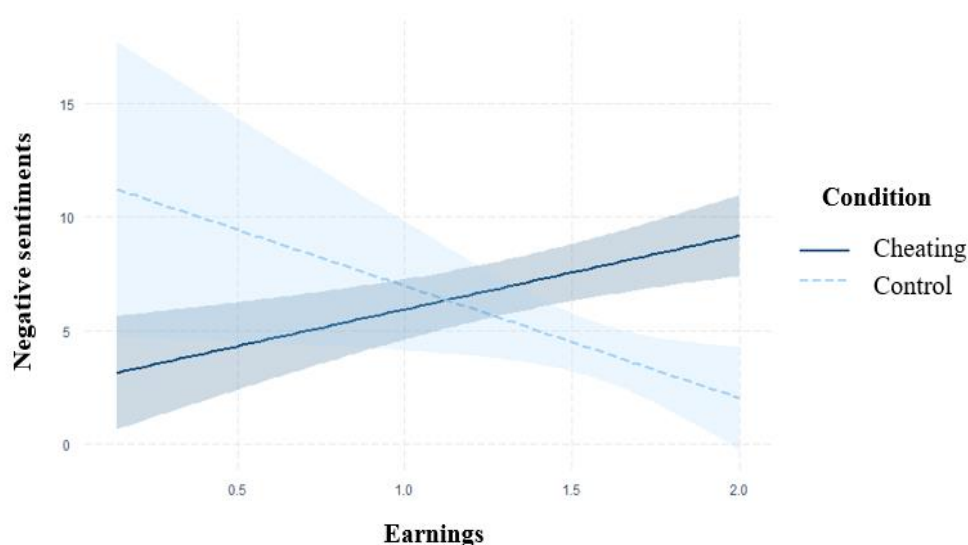
Figure 3.1. The difference in sentiments between cheating and control conditions



According to Tausczik & Pennebaker (2010), people lying or behaving unethically tend to use a higher total word count. I did not find a significant difference in the word count of the reviews between participants in the control group ($M = 20.73$, $SD = 15.61$, $Mdn = 17.50$) and the cheating group ($M = 20.03$, $SD = 16.40$, $Mdn = 15$; $t(199) = .31$, $p = .76$). However, a linear regression analysis revealed that in the cheating condition, there was a significant positive correlation between the level of earnings (represent the degree of cheating) and word count, $b = .20$, $F(1, 101) = 4.22$, $p < .05$. In contrast, the correlation between the level of earnings and word count in the control group was not significant, $b = -.02$, $F(1, 96) = .03$, $p = .87$.

To test hypothesis H3.1 (i.e., the correlation between earnings and sentiments), I used the PROCESS method (Model 1 in Hayes, 2017) with the condition as a moderator for the effect of earnings on negative sentiments. The analysis revealed a significant moderation effect of condition (cheating, control) on the correlation between earnings and negative sentiments, $F = 6.22$, $p < 0.01$, with positive coefficient for the interaction ($b = 8.21$, $SE = 2.92$, $p < .01$). As Figure 3.2 illustrates, there was a significant positive correlation between earnings and negative sentiments in the cheating condition, $b = .25$, $SE = 1.27$, $p < .05$, while in the control condition, there was a significant negative correlation, $b = -.21$, $SE = 2.31$, $p < .05$. In other words, participants in the cheating condition reviewed the app less favorably the more they cheated (using more negative sentiments). In contrast, participants in the control condition used fewer negative sentiments the more they earned from it.

Figure 3.2. The impact of earning (US\$) on negative sentiments in cheating and control conditions

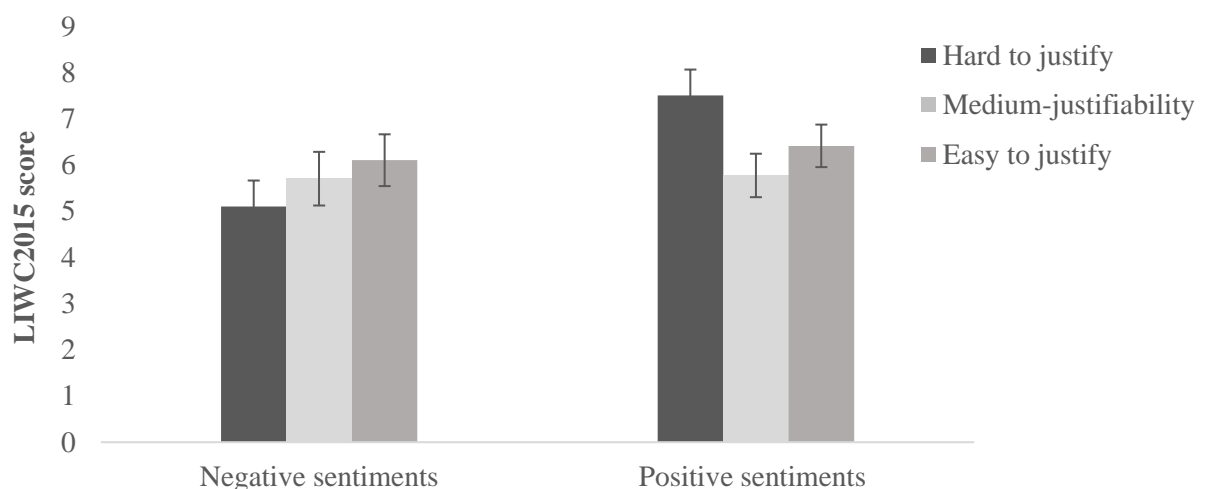


However, I found no support to hypothesis H3.2. A mediation analysis using PROCESS method (Model 4 in Hayes, 2017), revealed a non-significant indirect effect of condition (cheating and control) on negative sentiments score through earnings (the mediator), $b = -.07$, 95% $CI [-.17, .01]$. The lack of mediation effect is consistent with the findings of study 2.1 from which the data were taken and can attribute to the small sample.

Surprisingly, there was no significant correlation between earnings and positive sentiments ($F(1,199) = .89, p = .34$), in both control condition ($b = .10, SE = 4.46, p = .32$) and cheating condition ($b = -.04, SE = 1.13, p = .70$). The findings support the distancing account in which people judge more harshly the victims of their unethical behaviors. Considered jointly, the findings suggest that cheating was associated with harsher judgment of the app. These findings support the distancing account (Shalvi et al., 2015).

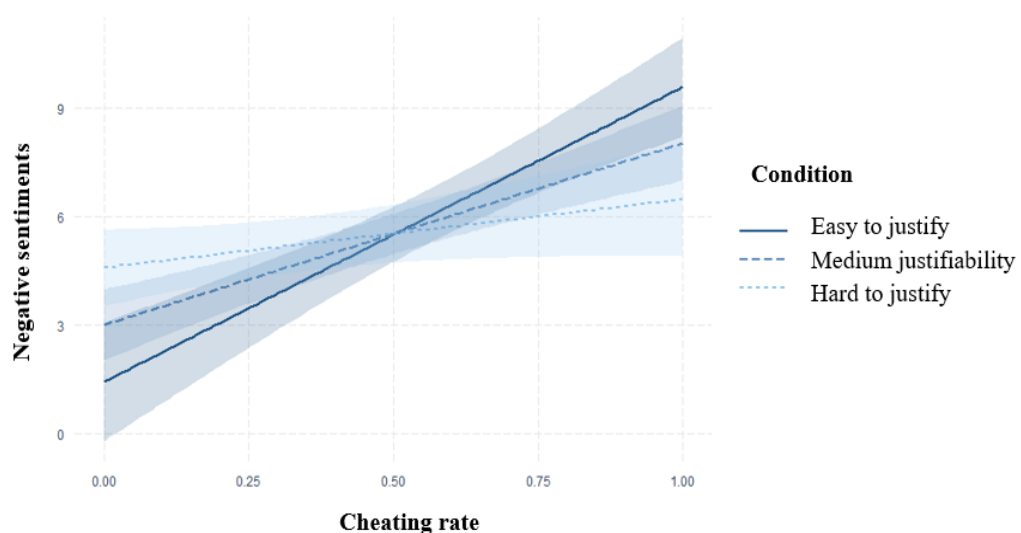
The ability to justify unethical behavior. According to the LIWC software, there was a significant difference in the positive sentiments between the hard-to-justify condition ($M = 7.50, SD = 6.90$), the medium-justifiability ($M = 5.77, SD = 5.80$) and the easy-to-justify condition ($M = 6.41, SD = 5.67$; $F(2,453) = 3.09, p < .05$). However, there was no significant difference in the negative sentiments according to conditions ($F(2,453) = 0.72, p = .49$) (Figure 3.3). According to MonkeyLearn software, there was a difference in the general sentiment tone. While the hard-to-justify general sentiment tone is 72.2% positive, the sentiment tone of the medium-justifiability condition and easy-to-justify conditions was negative (58.30% and 69.10%).

Figure 3.3. The difference in sentiments between cheating and control conditions



To test the correlation between cheating and sentiments, I used the PROCESS method (Model 1 in Hayes, 2017), with the condition as a moderator for the effect of cheating rate on negative sentiment. Consistent with hypothesis H3.3, the analysis revealed a significant moderation effect of condition (hard-to-justify, medium-justifiability, easy-to-justify) on the correlation between cheating and negative sentiments, $F = 10.1$, $p < 0.01$, with positive coefficient for the interaction ($b = 3.83$, $SE = 1.23$, $p < .01$). As Figure 3.4 illustrates, there was a significant positive correlation between cheating and negative sentiments in the easy-to-justify condition, $b = .33$, $SE = 2.01$, $p < .01$, and in the medium-justifiability condition, $b = .28$, $SE = 1.69$, $p < .01$. Meaning, the more participants cheat, the more they use negative sentiments to express harsh judgment. However, there was no significant correlation between cheating and negative sentiments in the hard-to-justify condition, $b = .03$, $SE = 1.58$, $p = .73$.

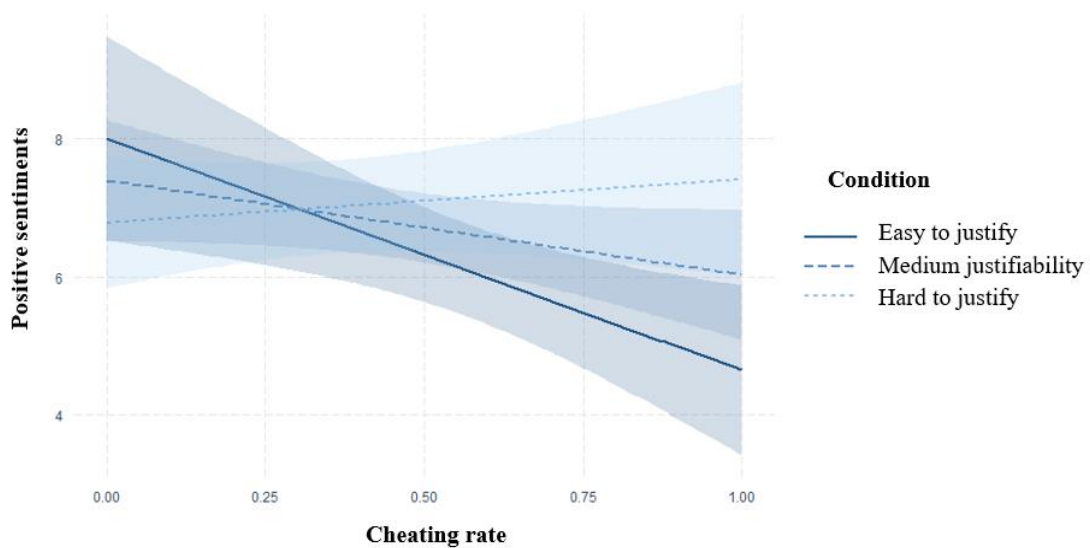
Figure 3.4. The impact of the cheating rate on negative sentiments by ease of justification condition



In order to test hypothesis H3.4 and estimate the direct effect of the condition (ease of justification) on negative sentiments score, and the indirect effect of the cheating rate as mediator, I performed a mediation analysis using PROCESS method (Model 4 in Hayes, 2017). Results indicated that condition was a significant predictor of cheating rate, $b = .31$, $SE = .02$, 95% CI [.09, .16], $p < .01$, and that cheating rate was a significant predictor of negative sentiments score, $b = .21$, $SE = 1.01$, 95% CI [2.38, 6.34], $p < .01$. Condition was no longer a significant predictor of negative sentiments score after controlling for the mediator, cheating rate, $b = -.01$, $SE = .41$, 95% CI [-.90, .73], $p = .84$, consistent with complete mediation.

Further examination revealed that there was a moderation effect of condition (hard-to-justify, medium-justifiability, easy-to-justify) on the correlation between cheating and positive sentiments, $F = 2.77, p < 0.05$, with negative coefficient for the interaction ($b = -2.44, SE = 1.11, p < .05$). As Figure 3.5 illustrates, and consistent with hypothesis H3.5, there was a significant negative correlation between cheating and positive sentiments in the easy-to-justify condition, $b = -.19, SE = 1.7, p < .05$. Meaning, the more participants cheat the less they express positive sentiments in their reviews. However, there was no significant correlation between cheating and positive sentiments in the medium-justifiability condition, $b = -.15, SE = 1.42, p = 0.07$, and in the hard-to-justify condition, $b = .11, SE = 1.57, p = .19$.

Figure 3.5. The impact of the cheating rate on positive sentiments by ease of justification condition



However, I found no support for hypothesis H3.6. A mediation analysis using PROCESS method (Model 4 in Hayes, 2017), revealed a non-significant indirect effect of condition on negative sentiments score through earnings (mediator), $b = -.12, 95\% CI [-.38, .11]$.

These results suggest that when it was easy-to-justify cheating, participants who cheated more tended to judge their victim more harshly by expressing more negative sentiments in their written reviews. However, there was no significant correlation in the hard-to-justify condition. The findings were conclusive only in the case of negative sentiments (hypothesis H3.4) and not positive sentiments (hypothesis H3.6). Importantly, the results support the distancing account because a harsher judgment was found only in conditions where participants could justify the unethical behavior and distance themselves from the victim. However, when

cheating was obvious, cheaters found it more difficult to distance themselves from the victim and therefore did not review the app more negatively.

Considered jointly, these findings suggest that unethical behavior impact not only the subsequent star rating of the target of transgression but also written reviews about the product and the experience while using it. Negative sentiments that represent a harsh judgment were more prominent when participants cheated more (hypotheses H3.1 and H3.3). Also, the results reinforce the distancing-based process as the underlying mechanism driving people's tendency to judge their target harshly. Meaning, the effect was found only when it was possible to justify the unethical behavior (easy to justify condition and medium justifiability). Interestingly, when it was easy to justify the misbehavior, there was also a negative correlation between unethical behavior and positive emotions (hypothesis H3.5), which implies that reviews were more consistent in terms of their sentiments. Consistency in sentiments may send cues to the reader regarding the review's accuracy and how helpful it is (Moore & Lafreniere, 2020; Schlosser, 2011). However, while the mediation role of the cheating rate on the effect of condition on negative sentiments was supported (hypothesis H3.4), it was not significant in positive sentiments (hypothesis H3.6). A possible explanation is that the distancing responses, which include negative judgment and blaming the victim of the transgression (Bandura, 1999), is usually expressed in more negative terms and sentiments rather than using less positive terms.

Online written reviews are widespread in social media, internet platforms, and blogs (Chuang, 2020), and people perceive them as a source of information. Considering that negative sentiment can jeopardize people's emotional trust, brand equity, and purchasing intentions (Hsu et al., 2017; Varga & Albuquerque, 2019), findings may have far-reaching implications for organizations, workplaces, and brands. For example, it seems that, paradoxically, firms that use lenient policies that put less emphasis on preventing unethical behavior may expose themselves to more negative online reviews.

However, according to recent studies, the influence of negative reviews in the online context may be limited since readers may suspect they were generated by competitors or represent different preferences (Babić et al., 2016). In addition, people tend to rely not only on sentiments while considering a product but also on other linguistic cues that may influence how a review is considered credible and trustworthy (Chen & Yuan, 2020). Thus, the next study explores whether the willingness to use a product is influenced by sentiments generated by people who cheated or behaved ethically.

Study 3.2 - The effect of reviews from ethical/unethical consumers on the willingness to use the product

The second study aimed to reveal if the differences in the review's sentiments (whether positive or negative) and the source of the review impact potential customers' willingness to use the product. Since the influence of negative written reviews in the online context is limited (Babić et al., 2016), the impact of unethical behavior found in Study 3.1 may not be significant to organizations and brands. It may be that negative sentiments are attributed to different apps' preferences, the internet connection that influences the experience, or other external factors.

In addition, positive and negative reviews can be written by both people that cheated or not. It is unclear if reviews written by people who cheated differ from those written by people that did not cheat. Studies show that, besides sentiments, people tend to pay attention to different linguistic cues in the reviews, such as words being used or the length of the review (Moore & Lafreniere, 2020). For example, consistent with studies suggesting that people lying or behaving unethically tend to use a higher total word count (Tausczik & Pennebaker, 2010), I found in Study 3.1, a positive correlation between cheating rate and word count (the correlation in the control group, where people did not cheat, was not significant). Therefore, it is not necessarily that positive or negative reviews will be perceived in the same way if they were generated by people who cheated or behaved ethically. The findings may be insightful since they may indicate that people who cheat use cues that convey a trustworthy and persuasive negative review (Chen & Yuan, 2020). It may also enable a way to identify reviews written by cheaters based on linguistic cues that characterize people who behave unethically.

In the study, I examine the willingness of participants to use a product (play the app, download the app, or pay for the app) after reading written reviews that vary according to whether the reviewers could cheat (control, cheating conditions), earnings level (low and high), and sentiments (positive or negative). I hypothesize that:

H3.7 People will be significantly less willing to use a product after reading a review written by a previous participant who behaved unethically compared to a review written by a previous participant who did not behave unethically.

H3.8 People will be significantly less willing to use a product after reading negative review written by a previous participant who behaved unethically compared to a negative review written by a previous participant who behaved ethically.





Method

Participants. I recruited 149 participants from Prolific (51% male, $M_{\text{age}} = 32$, $SD = 11.2$). All participants were above 18 years old, with English as their first language. Each participant was paid £0.40. The study was similar to all participants, and therefore had a minimum of 100 participants.

Design and Procedure. Participants were asked to imagine themselves in a situation in which they are looking for a new app, and during their search, they are reading few online reviews. Participants read a total of ten different written reviews and were asked, after each review, to provide their first impression and personal opinion on each.

The ten different reviews were chosen from all the reviews participants wrote at the dots task studies (Studies 2.1 and 2.2) and analyzed in Study 3.1. Each review represented a different condition, according to the following criteria: condition (cheating or control), earning (low or high. in the cheating condition, the earnings represent the level of cheating), sentiment according to LIWC score (positive or negative). Also, two reviews were taken from participants that had different ability to justify the cheating (easy-to-justify or hard-to justify) with a high level of cheating. The actual reviews for this study were randomly selected from each category (Figure 3.6) and presented to participants in random order. The reviews presented in the study included an indication of time and a unisex name to simulate real online reviews.

Figure 3.6. Images of reviews according to condition, earnings, and sentiments

Control condition	Cheating condition
Low earnings - Positive sentiments	
 Billie H. 2 months ago Game ran smoothly. Was entertaining, if a bit simple?	 Charlie D. 2 months ago It's cool yet challenging. The app would be good for someone who loves counting and fast paced games.
Low earnings - Negative sentiments	
 Blake S. 3 months ago The game was fun, but at the end it said I did not answer 10 times which is not true.	 Dylan A. 2 months ago I believe two seconds is too fast. I'm sure I answered but it said I didn't. I would give 3 seconds instead of two.

High earnings - Positive sentiments



Elliot L.

3 months ago

It was pretty fun trying to quickly guess the right answer! Some were a definite struggle, but it was challenging and enjoyable!



Alex Y.

3 months ago

Seems like a challenging game. Not sure how well you can improve on your performance.

High earnings - Negative sentiments



Taylor B.

3 months ago

The app was way too challenging for what I would want to do in my leisure time.



Kelsey S.

3 months ago

I don't really get it. I think I'd get bored very quickly just figuring out which side has more dots. It's not very much fun and I can't see how it would improve my skills in any way.

High earnings (cheating condition)



Easy to justify

Mickey M.

2 months ago

It is very difficult to guess which side has the most dots in 2 seconds the way the dots are randomly placed on the screen. Several times it did not register my selection when I pressed on my choice. I would not be very anxious to play this game just for the fun of it.



Hard to justify

Robin K.

2 months ago

Very boring game. It works well though.

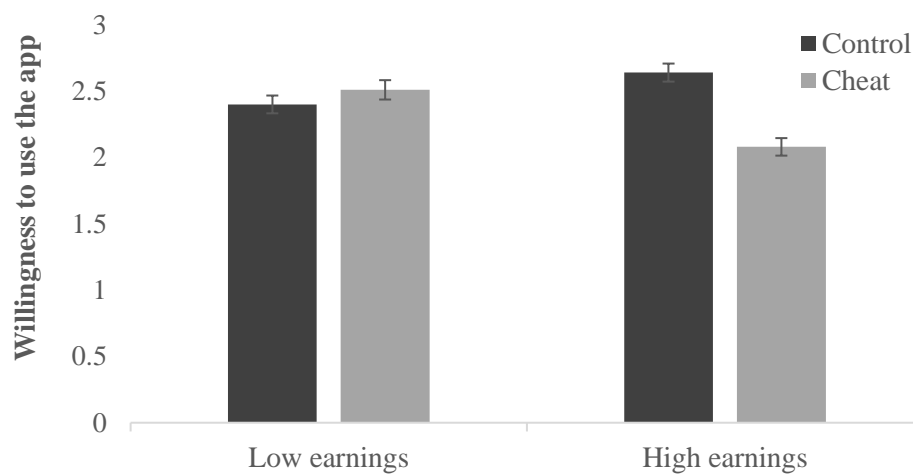
After each review, participants were requested to indicate whether they are willing to play the app, download it and pay an amount of £0.99 for it. Each of the measurements was tested on a 5-point-scale (1 – highly unlikely, 5 – highly likely). The three measures of their willingness to play, download, and purchase the app showed high internal reliability (Cronbach's $\alpha = .94$), so I averaged them to form a composite “general willingness to use the app” score. Lastly, participants reported their demographics and paid the promised amount when the study was completed.

Results and Discussion.

Consistent with hypothesis H3.7, I found that the willingness of participants to use the app after reading the reviews from the cheating condition was significantly lower compared to their willingness to use the app after reading reviews from the control condition ($M = 2.13, 2.52; SD = .79, .75; t(148) = 13.06, p < .01$). However, participants' willingness to use the app

differed significantly based on the level of earnings (which represent the degree of cheating in the cheating condition) between both conditions. The willingness of participants to use the app after reading reviews with low earnings was significantly higher in the cheating condition ($M = 2.51$, $SD = .89$) compared to the control condition ($M = 2.40$, $SD = .82$; $t(148) = 1.97$, $p < .05$). Yet, the willingness to use the app was reversed in the high earnings level, such that it is significantly lower after reading reviews from the cheating condition ($M = 2.08$, $SD = .81$) compared to reviews from the control condition ($M = 2.64$, $SD = .84$; $t(148) = 10.77$, $p < .01$). (Figure 3.7).

Figure 3.7. The willingness of participants to use the app after reading reviews according to the condition and earnings.

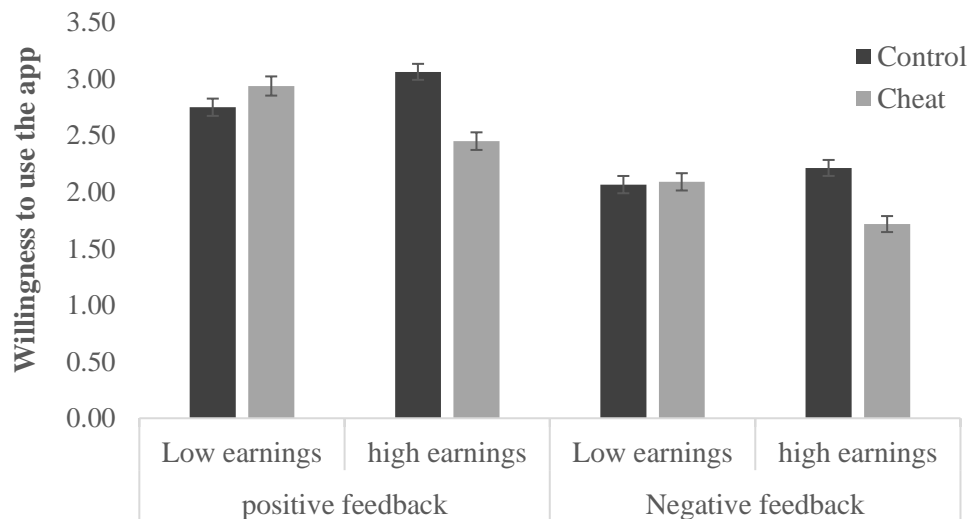


The same tendency was also found when participants read a positive review of the app. Participants were significantly more inclined to use the app after reading positive reviews from the cheating group with low earnings (meaning, low degree of cheating) compared to positive reviews from the control group with low earnings ($M = 2.93$, 2.75 ; $SD = 1.03$, $.91$; $t(148) = 2.41$, $p < .05$). However, participants were less inclined to use the app after reading positive reviews from the cheating group with high earnings (meaning, a high degree of cheating) compared to positive reviews from the control group with high earnings ($M = 2.45$, 3.06 ; $SD = .95$, $.91$; $t(148) = 8.67$, $p < .01$). (Figure 3.8).

In addition, consistent with hypothesis H3.8, the willingness of participants to use the app after reading negative reviews was not affected by the condition, whether cheating or control group, with low earnings ($M = 2.09$, 2.07 ; $SD = .93$, $.97$; $t(148) = .35$, $p = .73$). However, there was a significant difference in the willingness of participants to use the app after reading negative reviews from the cheating group with high earnings (i.e., a high degree of cheating)

compared to negative reviews from the control group with high earnings ($M = 1.72, 2.21; SD = .86, 1.01; t(148) = 7.09, p < .01$).

Figure 3.8. Participants' willingness to use the app after reading positive or negative reviews, according to the condition and earnings.



Surprisingly, there were no significant differences in the willingness of participants to use the app after reading negative reviews in the easy-to-justify with high degree of cheating ($M = 1.82, SD = .99$) and in the hard-to-justify with high degree of cheating ($M = 1.73, SD = .97; t(148) = 1.34, p = .18$). Meaning, although I found in Study 2.2 that participants in the hard-to-justify condition did not rank the app more negatively when cheating more, their negative review still resulted in a lower willingness to use the app.

Considered jointly, the findings suggest that reviews that were written after behaving unethically evoked less willingness of potential customers to use the product. Although negative reviews, whether generated after behaving ethically or unethically, lead to lower willingness to use the app in general, the reviews written after high cheating rates were significantly less attractive than negative reviews written by participants who did not cheat. Interestingly, after high cheating rates, even positive sentiments elicited a lower willingness to use the app than positive reviews written after gaining high earnings ethically.

Notably, reviews generated by participants with low levels of cheating increased people's willingness to use the product compared to the control group overall, and when reviews included positive sentiments. However, there was no difference in the willingness to use the app when the review included negative sentiments after low cheating rates compared to low earnings in the control group. Meaning, when people have the opportunity to cheat, but

despite the temptation, curb their own dishonesty (Ayal, Gino, Barkan, & Ariely, 2015), they tend to generate reviews that increase the willingness of potential customers to use the product at the best, or similar to people who did not have the opportunity to cheat.

The results reinforce the importance of considering unethical behavior and its impact on individual's judgment, word of mouth, and future decisions of others. It also emphasizes the importance of policies that aim to deter people from behaving unethically, not only in the marketplace but in different contexts in daily life. The implications are particularly relevant when considering the increasing use of social media as a trusted source for information and decision-making (Berger et al., 2020).

General Discussion

In this research, I examined how people judge those they have cheated, and more specifically, how their written reviews are affected by their unethical behaviors. In the first study, I show that when people behave unethically, they tend to use more negative sentiments in their written reviews and harshly judge the victim (hypothesis H3.1). However, the results are not fully supported and, similar to study 2.1, the mediation effect is not significant (hypothesis H3.2). Also, reinforcing the boundary condition found in Study 2.2, I find that unethical acts do not influence judgments and sentiments when cheating is blatantly obvious (hypotheses H3.3 and H3.5). Findings that judgments are less favorable only when the misbehavior can be justified support the distancing-based process as the mechanism that drives the effect. Interestingly, while the mediation of the cheating rate on the effect of the ability to justify the misbehavior on negative sentiments was significant (hypothesis H3.4), it was not significant in the case of positive sentiments (hypothesis H3.6). The lack of mediation effect in the case of positive sentiments may indicate that the distancing responses are being expressed by more negative terms (such as boring, frustrating and pointless) rather than using fewer positive terms (such as nice, fun, exciting).

Furthermore, I find that reviews written after cheating reduce potential customer's willingness to engage with the target of the transgression (Study 3.2; hypotheses H3.7-H3.8). While previous research shows that reviews affect consumers' evaluation of a brand and purchasing decisions (Babić et al., 2016), this study suggests that the effect of cheating may be more profound than merely negative sentiments. When comparing reviews with similar sentiments (negative or positive), readers' willingness to use a product was lower when generated by participants that cheated. An explanation can be found in the linguistic cues, style,

length of the review, or even words being used (Moore & Lafreniere, 2020). It is also possible that the balance between general sentiments, harsher expression of emotions, and focus on technical aspects send different signals to the reader. Further research can explore the various elements that influence people's intentions while reading reviews besides sentiments.

The results suggest that unethical behavior's negative consequences are relevant not only in star rating but also in written reviews. Considering that the digital environment enables people to share their experiences, not only as consumers but also as employees and citizens, it may significantly impact people's attitude toward the firm and even jeopardize the emotional trust toward the organization (Chuang, 2020; Kauffmann et al., 2019). Surprisingly, reviews generated by people with low cheating levels included more positive sentiments than those written by reviewers who were not given the opportunity to cheat. These findings suggest that when people limit their own dishonesty (Ayal et al., 2015), they tend to provide more favorable reviews, presumably, since they feel good about their honest behavior, although the temptation.

The research findings have important implications for managers and marketers because of how people who behave unethically communicate and share their opinions about those they have cheated. In the marketing realm, subsequent reviews may impact purchasing decisions, brand equity, and consumer-brand relationships (Khamitov, Wang, & Thomson, 2019). Organizations and government entities may endure lower trust (Hsu et al., 2017) and possible damage to their image (Driss et al., 2019; Packard & Berger, 2017). As a result, organizations might find it challenging to recruit and maintain good employees, provide quality services, or keep the stakeholders' interests (Singh & Twalo, 2015).

In addition, findings suggest implications for policymakers and organizations regarding the importance of regulations that deter people from unethical behavior. Since Study 3.1 provide evidence for positive sentiments when people curb their dishonesty below the maximum possible (Mazar, Amir, & Ariely, 2008), it may be worthwhile for organizations and governments to invest in policies that minimize people's unethical behaviors while reinforcing their sense of moral decisions. Also, since dishonesty may impact not only star rating but customer-generated content in general, policymakers should consider this phenomenon while striving to maintain their image in the public eyes, improve the level of trust of citizens, and recruit employees that can, eventually, improve the services provided to citizens (Gordon, 2000; Hsu et al., 2017; Singh & Twalo, 2015).

Limitations and future research directions

The research expands the findings of previous studies in the dissertation (Chapters 1 and 2). It reinforces the notion that unethical behavior has a substantial impact on people's subsequent judgment and behavior. However, although the results suggest that dishonest people may distance themselves from the acts to restore their moral self-image, the phenomenon's extent is still unclear. People who behave unethically may choose not to write online reviews or only provide star ratings when possible. Future studies can explore whether people behaving unethically are willing to post written reviews in similar or different magnitude than people who behave ethically.

Future research may also explore the differences in the linguistic cues between written reviews generated after behaving ethically or unethically. As suggested in previous studies (e.g., Chen & Yuan, 2020), people consider not only sentiments in the review but also other cues such as style, narrative, length, and even the use of dispreferred markers (Moore & Lafreniere, 2020). As a result, and as evident in Study 3.2, the evaluation of reviews, similar in their sentiments, may be different when generated after cheating. Revealing possible linguistic cues can assist in identifying potential biased online reviews.

Lastly, while considering moderating factors, there may be differences in how they moderate the effect of unethical behavior on subsequent star ratings compared to the written review. For example, since reviews include positive and negative sentiments, it may be that transgression toward small companies with identifiable victims (like small ice-cream shops or café) will have a more significant impact on the tone or sentiments of the written reviews compared to reviews of large companies (Gino et al., 2010; Wirtz and McColl-Kennedy 2010). In addition, it is not clear if the time between the actual transgression and the review can impact the sentiments expressed in it and the words being used to describe the experience or product for better or worse.

Conclusion

In conclusion, this research shed light on the consequence of unethical behavior and how people judge those they have cheated, and more specifically, the tone and sentiments they use while writing their review or sharing their experience online. Behaving unethically may encourage people to distance themselves and harshly judge the company by providing negative online reviews. Due to the importance of customer-generated content and online media platforms, further harm can be caused to brands or organizations that were already cheated.

This research has practical implications for managers, marketers, and policymakers in minimizing the negative effects of unethical behavior in different settings of daily life. One of the issues that can be considered is whether policymakers and governmental regulations should intervene to reduce possible damages in the marketplace. Interventions may address both consumers (that may take decisions based on biased information) and organizations (that may face loss of revenue along with possible harm to their reputation or image in the eyes of the public). Additional studies on the impact of unethical behavior on subsequent judgment are needed in order to weigh the options and consider possible interventions (if any).

References

- Adamopoulos, P., Ghose, A., & Todri, V. (2018). The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research*, 29(3), 612-640.
- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. Mikulincer M, Shaver PR, eds. *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (American Psychological Association, Washington, DC) 149-166.
- Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science*, 10(6), 738-741.
- Babić, Rosario. A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318.
- Bambauer-Sachse, S., & Mangold, S. (2011). Brand equity dilution through negative online word-of-mouth communication. *Journal of retailing and consumer services*, 18(1), 38-45.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757-773.
- Basmmi, A. B. M. N., Abd Halim, S., & Saadon, N. A. (2020). Comparison of Web Services for Sentiment Analysis in Social Networking Sites. In *IOP Conference Series: Materials Science and Engineering* (Vol. 884, No. 1, p. 012063). IOP Publishing.
- Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of consumer psychology*, 24(4), 586-607.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, 84(1), 1-25.

- Bertot, J. C., Jaeger, P. T., & Hansen, D. (2012). The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government information quarterly*, 29(1), 30-40.
- Bhadane, C., Dalal, H., & Doshi, H. (2015). Sentiment analysis: Measuring opinions. *Procedia Computer Science*, 45, 808-814.
- Blackshaw, P., & Nazzaro, M. (2006). Word of mouth in the age of the web-fortified consumer. *Consumer-generated media (CGM)*, 101.
- Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6), 953-975.
- Chen, Y. J., & Kirmani, A. (2015). Posting strategically: The consumer as an online media planner. *Journal of Consumer Psychology*, 25(4), 609-621.
- Chen, Z., & Yuan, M. (2020). Psychology of word of mouth marketing. *Current opinion in psychology*, 31, 7-10.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Chuang, S. H. (2020). Co-creating social media agility to build strong customer-firm relationships. *Industrial Marketing Management*, 84, 202-211.
- Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of personality and social psychology*, 100(5), 947-966.
- Driss, O. B., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly*, 36(3), 560-570.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research*, 19(3), 291-313.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1-44
- Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological science*, 21(5), 712-720.

- Gneezy, U., Imas, A. & Madarász, K. (2012). Conscience accounting: emotional dynamics and social behaviour. *Theoretical economics paper series*, STICERD, London, UK.
- Gordon, M. T. (2000). Public trust in government: The US media as an agent of accountability?. *International Review of Administrative Sciences*, 66(2), 297-310.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- Hochman, G., Glöckner, A., Fiedler, S., & Ayal, S. (2016). “I can see it in your eyes”: Biased Processing and Increased Arousal in Dishonest Responses. *Journal of Behavioral Decision Making*, 29(2-3), 322-335.
- Houston, D. J., Aitalieva, N. R., Morelock, A. L., & Shults, C. A. (2016). Citizen trust in civil servants: A cross-national examination. *International Journal of Public Administration*, 39(14), 1203-1214.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision support systems*, 52(3), 674-684.
- Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274-1306.
- Hsu, C. L., Yu, L. C., & Chang, K. C. (2017). Exploring the effects of online customer reviews, regulatory focus, and product type on purchase intention: Perceived justice as a moderator. *Computers in Human Behavior*, 69, 335-346.
- Im, T., Cho, W., Porumbescu, G., & Park, J. (2014). Internet, trust in government, and citizen compliance. *Journal of Public Administration Research and Theory*, 24(3), 741-763.
- Ismagilova, E., Slade, E. L., Rana, N. P., & Dwivedi, Y. K. (2019). The effect of electronic word of mouth communications on intention to buy: a meta-analysis. *Information Systems Frontiers*, 1-24.
- Ivanova, O., & Scholz, M. (2017). How can online marketplaces reduce rating manipulation? A new approach on dynamic aggregation of online ratings. *Decision Support Systems*, 104, 64-78.
- Jia, S. (2018). Behind the ratings: Text mining of restaurant customers’ online reviews. *International Journal of Market Research*, 60(6), 561-572.

- Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American journal of psychology*, 263-286.
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*.
- Khamitov, M., Wang, X., & Thomson, M. (2019). How well do consumer-brand relationships drive customer brand loyalty? Generalizations from a meta-analysis of brand relationship elasticities. *Journal of Consumer Research*, 46(3), 435-459.
- Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing*, 33(1), 11-26.
- Kouchaki, M., & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological science*, 25(1), 95-102.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands?. *Management Research Review*.
- Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial marketing management*, 90, 493-507.
- Lee, M., & Youn, S. (2009). Electronic word of mouth (eWOM) How eWOM platforms influence consumer product judgement. *International Journal of Advertising*, 28(3), 473-499.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6), 633-644.
- Moore, S. G., & Lafreniere, K. C. (2020). How online word-of-mouth impacts receivers. *Consumer Psychology Review*, 3(1), 34-59.
- Murphy, R. (2019). *Local Consumer Review Survey: Online Reviews Statistics & Trends*. Retrieved from <https://www.brightlocal.com/research/local-consumer-review-survey-2019/>

- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.
- Nielsen (2015). *Global trust in advertising winning strategies for an evolving media landscape*. Retrieved from <https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/global-trust-in-advertising-report-sept-2015-1.pdf>
- Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research*, 54(4), 572-588.
- Pang, B., Lee, I. (2008). Opinion Mining and Sentiment Analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Pennebaker, J. W., Booth, R., & Francis, M. E. (2007). Linguistic Inquiry and Word Count (LIWC2007).
- Rahman, K. M., & Khamparia, A. (2016). Techniques, applications and challenges of opinion mining. *IJCTA*, 9(41), 455-461.
- Reimer, T., & Benkenstein, M. (2016). When good WOM hurts and bad WOM gains: The effect of untrustworthy online reviews. *Journal of Business Research*, 69(12), 5993-6001.
- Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology*, 21(3), 226-239.
- Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of interactive marketing*, 21(4), 76-94.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125-130.
- Simonson, I., & Rosen, E. (2014). What marketers misunderstand about online reviews. *Harvard Business Review*, 92(1), 7.
- Singh, P., & Twalo, T. (2015). Mismanaging unethical behaviour in the workplace. *Journal of Applied Business Research (JABR)*, 31(2), 515-530.

- Tata, S. V., Prashar, S., & Gupta, S. (2020). An examination of the role of review valence and review source in varying consumption contexts on purchase decision. *Journal of Retailing and Consumer Services*, 52.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.
- The CMO survey (2020). *Highlights and Insights Report, February 2020*. Retrieved from https://cmosurvey.org/wp-content/uploads/2020/02/The_CMO_Survey-Highlights-and_Insights_Report-Feb-2020.pdf
- Tracy, J. L., & Robins, R. W. (2007). The self in self-conscious emotions: A cognitive appraisal approach. In J.L. Tracy, R. W. Robins, & J.P. Tangney (Eds.) *The self-conscious emotions: Theory and research*, (pp. 3-20). New York: Guilford Press.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of marketing*, 73(5), 90-102.
- Varga, M., & Albuquerque, P. (2019). Measuring the impact of a single negative consumer review on online search and purchase decisions through a quasi-natural experiment. *Marketing Science Institute*.
- Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, 43(6), 875-894.
- Wirtz, J., & McColl-Kennedy, J.R. (2010). Opportunistic customer claiming during service recovery. *Journal of the Academy of Marketing Science*, 38(5), 654-675.
- Zhang, K. Z., Zhao, S. J., Cheung, C. M., & Lee, M. K. (2014). Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model. *Decision Support Systems*, 67, 78-89.
- Zhong, N., & Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science*.

Ziegele, M., & Weber, M. (2015). Example, please! Comparing the effects of single customer reviews and aggregate review scores on online shoppers' product evaluations. *Journal of Consumer Behaviour*, 14(2), 103-114.

Discussion and Conclusions

Although unethical behavior is a well-known phenomenon (e.g., Gino, 2015; Shalvi, Eldar, & Bereby-Meyer, 2012), much less is known about its impact on subsequent judgments and evaluations of the victim. Subsequent judgments have the potential to carry additional implications to organizations, brands, and government authorities and affect their performance and reputation. Critically, current theories provide conflicting predictions for the possible impact of the dishonesty of people's judgments and evaluations of the target of the transgression. While some studies predict higher and more positive subsequent judgments following dishonesty (i.e., guilt account; Cohen, Wolf, Panter, & Insko, 2011; Sachdeva, Iliev, & Medin, 2009; Xu, Bègue, & Bushman, 2012), other studies predict lower and more negative subsequent judgments following unethical behaviors (e.g., distancing account; Ayal & Gino, 2011; Barkan, Ayal, Gino, & Ariely, 2012; Shalvi, Gino, Barkan, & Ayal, 2015).

In the digital era, the ease of sharing information and opinions by customers, employees, and citizens is apparent (Chuang, 2020; Kauffmann, Peral, Gil, Ferrández, Sellers, & Mora, 2019). Judgments and evaluations of people are often shared online and can reach thousands of people (Babić, Sotgiu, De Valck, & Bijmolt, 2016), with an enormous effect on business and government organizations' performance. Hence, the possibility that unethical behavior may have lingering effects, besides the immediate financial costs that are often integrated into the profit and loss reports (mostly by increasing the price of consumer goods; Feldman, van Rooij, & Rorie, 2019), is an important factor that needs to be considered while formulating business strategies or policies. Brand equity, firms' images, organization reputation, citizens trust, and the ability to implement public policies are few examples of the possible effect of unethical behavior (Driss, Mellouli, & Trabelsi, 2019; Gordon, 2000; Hsu, Yu, & Chang, 2017; Khamitov, Wang, & Thomson, 2019; Packard & Berger, 2017; Singh & Twalo, 2015). The aim of my research is to explore the phenomenon and reveal its possible implications.

The first objective of the current research was to contrast these opposing predictions to better understand how people judge those they cheat. Subsequent judgment of the target of the transgression was examined in different settings: organizations, government authorities, and the marketing realm, with both real-life cases and online performance. Across studies, I found that people tend to judge those they have cheated more negatively, and the more they

cheat, the more negative these judgments become. The results were found among employees who recalled their unethical behaviors in the workplace (Studies 1.1–1.3), self-employed who recalled their false tax reports (Study 1.4), and in a marketing setting with a new app (Study 2.1). However, while the findings imply that people tend to distance themselves by judging the target of their transgression in an attempt to justify their actions (Barkan et al., 2012), some inconsistencies in the results need to be addressed. Further studies are needed to examine the phenomenon among self-employed (by using a larger sample or asking participants to write questionable behaviors instead of receiving a list of behaviors) and additional marketing settings (or tasks) to confirm the consistency of the phenomenon.

Notably, the findings indicate that this negative consequence of unethical acts is likely to be quite pervasive because it happens in both long-term relationships, such as with an employer, but also in one-time interactions, such as when using a particular product or app. In addition, using the recall paradigm (Fox & Kahneman, 1992), I was able to refute a reverse-causality account whereby people judge the victim negatively and therefore are more willing to cheat (Studies 1.3, 1.4). Meaning, contrary to previous studies (e.g., Dell’Anno, 2009) and common beliefs among the general public, it is possible that unethical behaviors, such as tax evasion, drive low satisfaction and not the opposite.

The second objective of the current research was to explore the moderating role of justification and understand whether subsequent judgment is affected by the ability to justify unethical behavior. Demonstrating an important boundary condition, I found that unethical behavior does not influence judgments when cheating is blatantly obvious and hard to justify (Study 2.2). This finding supports the distancing-based process as the underlying mechanism driving the effect. Meaning, paradoxically, organizations that use lenient policies or not preventing (or limiting) unethical behavior may expose themselves to more negative subsequent judgments.

In addition, I examined two additional moderating factors. First, I found support for the moderating role of moral credentials (Study 2.3), in which people tend to cheat more after behaving morally (Effron, Cameron, & Moni, 2009). Findings suggest that moral credentials (after earning money ethically) provided the ability to justify cheating, even when it was blatantly obvious (i.e., hard to justify), and to use a distancing account as a post justification mechanism. However, I found no evidence for the moderating role of actively benefit from cheating (Muncy & Vitell, 1992). Findings suggest that while cheating rates are higher when

passively benefiting on others' expense, there was no significant difference in the subsequent rating whether the benefit was acquired actively or passively (Study 2.5).

The third objective of the current research was to explore the impact of unethical behavior on written review's nature and sentiments and the possible influence they have on other people's behaviors and willingness to engage with the victim. Importantly, written reviews have unique characters that distinguish them from star-rating; they include negative, positive, or mixed sentiments not always consistent with the star rating (Chen & Yuan, 2020), they provide room to express opinions (Jia, 2018), can be found in social media, blogs, and internet forums separately from ranking platforms (Chuang, 2020), and may influence people emotional trust (Hsu et al., 2017).

I found that when people behave unethically, they tend to use more negative sentiments in their written reviews and harshly judge the victim (Study 3.1). Meaning, unethical behavior impacts the subsequent star rating of the target of transgression and also written reviews about a product or a service. Also, reinforcing the boundary condition revealed in Study 2.2, I found that unethical acts do not influence judgments and sentiments when cheating is blatantly obvious (Study 3.1). Findings that judgments are less favorably only when the misbehavior can be justified support the distancing-based process as the underlying mechanism driving people's tendency to judge their target harshly.

Furthermore, while Studies suggest that negative sentiments may be limited in the online context (Babić et al., 2016), I found that reviews written after cheating significantly reduced readers' willingness to engage with the target of the transgression (Study 3.2). Notably, the findings suggest that the effect of cheating may be more profound than merely negative sentiments. When comparing reviews with similar sentiments (negative or positive), readers' willingness to use a product was significantly lower when generated by participants that cheated. An explanation can be found in the linguistic cues (e.g., style, length of the review), which increase the written review's credibility (Chen & Yuan, 2020; Moore & Lafreniere, 2020).

Surprisingly, reviews generated by participants who had the opportunity to cheat, but despite the temptation, curb their own dishonesty (Ayal, Gino, Barkan, & Ariely, 2015) increased readers' willingness to use the product compared to those who behaved ethically. A possible explanation is that people feel good about their moral decision when facing temptation and therefore tend to provide more favorable reviews. This finding suggests that it may be

worthwhile for firms to develop policies that aim to minimize people's unethical behavior while reinforcing (or acknowledging) their moral decisions.

Theoretical and practical contribution

The research has a significant theoretical contribution. Examining the impact of unethical behavior on subsequent judgments of the victim is innovative since it highlights an unexplored aspect of unethical behavior. The research considers the impact unethical behavior has on subsequent judgments and evaluations rather than focus on antecedents and ways to mitigate it. Also, findings add to the extant literature by considering not only the direct costs of unethical behavior (e.g., lost revenues) but also its lingering effects on judgments, word of mouth, and consequently other people's behaviors.

Additionally, the research contributes both to the theory and practice of online reviews (Babić et al., 2016; Chen & Kirmani, 2015; Chevalier & Mayzlin, 2006) by revealing a novel factor that can systematically influence ratings and written reviews. Given that unethical behavior is prevalent, with up to 64% of consumers misbehaving at some point in their lives (Farmer & Dawson, 2017), one out of every 50 employees caught stealing from their workplace (Hayes international annual retail theft survey, 2019), and \$406 billion in lost funding annually in the United States due to tax evasion (IRS, 2016), many online reviews and opinions likely originate from unethical behavior. The findings suggest that these reviews may be biased because people who behave unethically distance themselves from the act. However, the phenomenon's extent is unclear because people who behave unethically do not necessarily write or share a review. Future studies can explore whether people behaving unethically post online reviews more or less than their counterparts, and to what extent.

Furthermore, the research uses innovative methodological tools to explore the impact of unethical behavior on subsequent judgments. To avoid using only tasks in laboratory settings with low ecological validity (Ayal, Celse, & Hochman, 2019), I changed the dots task (Hochman, Glöckner, Fiedler, & Ayal, 2016) to a demo of an app and asked participants to test it and rate the different aspects of it. Playing the game online in a more familiar interface enabled me to better evaluate the subsequent judgment of the target of transgression. In addition, when exploring subsequent judgments in real-life situations (e.g., unethical behavior among employees), I used the recall paradigm (Fox & Kahneman, 1972), which manipulates the order of two questions to examine whether the evaluation of satisfaction (or judgment in

general) varies with the salience of another measure. Using this paradigm enabled to dismiss a reverse causality in which negative judgment allowed people to cheat more.

The research has important practical implications for organizations because of how unethical people can easily communicate with others about the organization they have cheated. In today's digital environment, organizations are subject to public opinion, and people can easily share their opinions on different platforms such as blogs, social media, and internet forums (Chuang, 2020). Thus, people who behave unethically can easily harm the victim of their transgression. As a result, the organization might find it challenging to maintain their image in the public's eyes, recruit and maintain good employees, provide quality services, or keep the stakeholders' interests (Singh & Twalo, 2015).

Furthermore, the research may have far-reaching implications for firms, policymakers, and marketers in terms of business strategies and effective policies. Findings emphasize the importance of policies that aim to deter people from behaving unethically due to the negative consequences to people's attitude and emotional trust toward organizations or brands (Chuang, 2020; Kauffmann et al., 2019). Based on Study 2.2, strict policies can influence the ability to justify unethical behavior. Meaning, when misbehavior is obvious and hard to justify, cheating rates decrease, and there is no impact of dishonesty on negative subsequent judgments of the target of the transgression. Lastly, firms can decide to shift part of their marketing resources to creative efforts along the customer journey to minimize the possible downstream consequence of dishonesty.

Lastly, the research suggests implications concerning consumer protection policies. Since unethical behavior has significant consequences in terms of word of mouth, people should be aware of possible biases that may influence their decision-making while considering a product or interacting with organizations. Additional studies are needed to reveal potential cues typical to online reviews following unethical behavior. Current studies explore possible cues of fake reviews by competitors or owners of companies by detecting textual markers of fakery or using detection algorithms (e.g., Barbado, Araque, & Iglesias, 2019; Mayzlin, Dover, & Chevalier, 2014) but do not address the possibility of biased reviews following unethical behavior. Revealing these cues will offer policymakers the possibility to monitor suspicious word of mouth through text analysis or machine learning software or use other methods based on this knowledge to protect the interests of customers, organizations, and entities that may be harmed due to biased online reviews.

Limitations and Future research directions

While the research suggests a distancing-based process as the underlying mechanism driving people's tendency to harshly judge the victims, a better understanding of the distancing mechanics is needed. It is unclear if people distance themselves to relieve ethical dissonance (Shalvi et al., 2015), use a moral self-regulation disengagement to reframe the act or attribution of blame to the victim (Bandura, 1999). In addition, while moral disengagement enables to avoid threats to the moral self-image, it is also possible that using moral disengagement strategies makes unethical behaviors personally acceptable (Shu, Gino, & Bazerman, 2011) and consequently reduces the need to justify the action by harsh judgment of the victim.

In addition, future research may uncover moderating factors. For example, the guilt account predicts that cheaters will rate the target of their transgression more favorably to compensate them and repent for the actions (Gneezy, Imas, & Madarász, 2012; Shu & Gino, 2012). While I found no evidence in my studies for the guilt account, additional studies can explore cases in which guilt moderate the subsequent judgment of the victim. Such cases can be, for example, transgression toward small identified companies (e.g., local grocery shop, café) versus large companies (Gino, Shu, & Bazerman, 2010; Wirtz and McColl-Kennedy 2010).

Another moderating factor can be individual differences in beliefs in a just world (BJW) moderate the subsequent judgment of the victim. According to the just-world theory, people need to believe the world is a predictable place where people get what they deserve (Lerner, 1980; Wilson & Darke, 2012). Thus, people with a high BJW may tend to judge the victim more negatively because they may believe the victim is deserving of antisocial behavior in the first place (Wenzel, Schindler, & Reinhard, 2017). In addition, I did not manipulate victim characteristics. Social equity theory posits that people may be less likely to cheat and subsequently negatively judge those who have treated them fairly (Adams, 1965). In a seminal study, Greenberg (1990) found that employees who received a thorough and sensitive explanation for a pay cut felt less inequality and stole less from their employer. Moreover, in an experimental setting, Houser, Vetter, & Winter (2012) found that participants who felt they were treated fairly as the allocation recipient in a dictator game were less likely to subsequently cheat than those who felt they were treated unfairly. Future research may examine whether organizations that generate goodwill and are perceived as more trustworthy are less subject to individuals' unethical acts and their downstream consequences.

Lastly, future research may explore differences in reviews' linguistic cues and customer-generated content after behaving ethically or unethically. Besides sentiments, reviews can include different linguistic cues such as style, narrative, length, and even the use of dispreferred markers (Chen & Yuan, 2020). These cues can impact the perceived credibility and level of the review's persuasiveness (Moore & Lafreniere, 2020). Revealing possible linguistic cues associated with unethical behavior can assist in identifying potential biased online reviews. Importantly, it is unclear whether the time between the actual transgression and the review can impact the sentiments expressed in it and the words being used to describe the experience or product for better or worse. Evidence shows that immoral choices need to be recent since unethical behavior tends to decay over time (Gneezy et al., 2012).

Conclusion

In conclusion, my research project has uncovered an unexplored consequence of unethical behavior, that is, how people judge those they have cheated. Behaving unethically may result in harsh judgment of the target of the transgression, presumably in an effort of people to distance themselves from the act and restore their moral self-image. Subsequent judgments and evaluations, whether in the form of rating, written reviews, recommendations, and online content, may influence organizations' image and, consequently, other people's behavior. In addition, the research suggests that the ability to justify unethical behavior serves as the underlying mechanism of the phenomenon, making it possible to blame the victim only when the misbehavior can be justified.

This research lays the foundation for future examination of a phenomenon that can affect organizations and ultimately on us as consumers and citizens. Despite data on the extent of unethical behavior in various settings, the damage to organizations due to subsequent judgments of unethical people appears to be profound, affecting reputation, public trust, the ability to provide good service, and implementing effective public policies. However, the lack of studies on the possible outcomes of unethical behavior and ways to address them has led to managerial and business strategies that did not consider these factors. Further understanding of the mechanism that influences subsequent judgments, possible moderating factors of the effect, and cues that enable to identify possible biased word-of-mouth that influence other people's decisions and behaviors are essential to business and government organizations and can have a significant effect on their performance, as well as on the welfare of individuals worldwide.

References

- ACFE (2020). *2020 Global study on occupational fraud and abuse*. Retrieved from <https://acfepublic.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf>
- Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, 2, 267-299.
- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, 83(6), 1423-1440.
- Ayal, S., Celse, J., & Hochman, G. (2019). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes*.
- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. In M. Mikulincer & P.R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 149-166). Washington, DC: American Psychological Association.
- Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science*, 10(6), 738-741.
- Ayal, S., Hochman, G., & Ariely, D. (2016). Dishonest behavior, from theory to practice. *Frontiers in Psychology*, 7, 1521.
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Barbado, R., Araque, O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. *Information Processing & Management*, 56(4), 1234-1244.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757-773.

- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169-217.
- Business Software Alliance (2018). *2018 BSA Global Software Piracy Study*. Retrieved from <https://gss.bsa.org/>.
- Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, 34(1), 1-25.
- Chen, Y. J., & Kirmani, A. (2015). Posting strategically: The consumer as an online media planner. *Journal of Consumer Psychology*, 25(4), 609-621.
- Chen, Z., & Yuan, M. (2020). Psychology of word of mouth marketing. *Current opinion in psychology*, 31, 7-10.
- Cheung, C. M., & Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision support systems*, 53(1), 218-225.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Chuang, S. H. (2020). Co-creating social media agility to build strong customer-firm relationships. *Industrial Marketing Management*, 84, 202-211.
- Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of personality and social psychology*, 100(5), 947-966.
- De Pelsmacker, P., Driesen, L., & Rayp, G. (2005). Do consumers care about ethics? Willingness to pay for fair-trade coffee. *Journal of consumer affairs*, 39(2), 363-385.
- Dell'Anno, R. (2009). Tax evasion, tax morale and policy maker's effectiveness. *The Journal of Socio-Economics*, 38(6), 988-997.
- Driss, O. B., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly*, 36(3), 560-570.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of experimental social psychology*, 45(3), 590-593.

- Farmer, J. F., & Dawson, J. (2017). American College Students' Shoplifting Experience: A Comparison of Retrospective Self-Reports to Micro-Level Criminological Theory. *International Journal of CriminFal Justice Sciences*, 12(1), 1.
- Feldman, Y., van Rooij, B., & Rorie, M. (2019). Rule-breaking without Crime: Insights from Behavioral Ethics for the Study of Everyday Deviancy. *Feldman, Y., Rorie, M., & Van Rooij, B.(2019). Rule-breaking without Crime: Insights from Behavioral Ethics for the Study of Everyday Deviancy. The Criminologist*, 44(2), 8-11.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525-547.
- Fox, C. R., & Kahneman, D. (1992). Correlations, causes and heuristics in surveys of life satisfaction. *Social Indicators Research*, 27(3), 221-234.
- Gamliel, E., & Peer, E. (2013). Explicit risk of getting caught does not affect unethical behavior. *Journal of Applied Social Psychology*, 43(6), 1281-1288.
- Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1-44.
- Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current opinion in behavioral sciences*, 3, 107-111.
- Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless+ harmless= blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes*, 111(2), 93-101.
- Gneezy, U., Imas, A. & Madarász, K. (2012). Conscience accounting: emotional dynamics and social behaviour. *Theoretical economics paper series*, STICERD, London, UK.
- Gordon, M. T. (2000). Public trust in government: The US media as an agent of accountability?. *International Review of Administrative Sciences*, 66(2), 297-310.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75(5), 561-568.
- Hayes international annual retail theft survey (2019). *32nd Annual retail theft survey*. Retrieved from <http://hayesinternational.com/news/annual-retail-theft-survey>

- Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, 49(2), 263-266.
- Hochman, G., Glöckner, A., Fiedler, S., & Ayala, S. (2016). "I can see it in your eyes": Biased Processing and Increased Arousal in Dishonest Responses. *Journal of Behavioral Decision Making*, 29(2-3), 322-335.
- Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, 345(6202), 1340-1343.
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.
- Hsu, C. L., Yu, L. C., & Chang, K. C. (2017). Exploring the effects of online customer reviews, regulatory focus, and product type on purchase intention: Perceived justice as a moderator. *Computers in Human Behavior*, 69, 335-346.
- Hunt, S. D., & Vitell, S. J. (2006). The general theory of marketing ethics: A revision and three questions. *Journal of Macromarketing*, 26(2), 143-153.
- Insurance Information Institute (2016). *Facts + Statistics: Fraud*. Retrieved from <http://www.iii.org/fact-statistic/fraud>
- IRS - Internal Revenue Service. (2016). *Tax Gap Estimates for Tax Years 2008-2010*. Retrieved from <https://www.irs.gov/newsroom/the-tax-gap>
- Jia, S. (2018). Behind the ratings: Text mining of restaurant customers' online reviews. *International Journal of Market Research*, 60(6), 561-572.
- Jussim, L., Yen, H., & Aiello, J. R. (1995). Self-consistency, self-enhancement, and accuracy in reactions to feedback. *Journal of experimental social psychology*, 31(4), 322-356.
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*.
- Khamitov, M., Wang, X., & Thomson, M. (2019). How well do consumer-brand relationships drive customer brand loyalty? Generalizations from a meta-analysis of brand relationship elasticities. *Journal of Consumer Research*, 46(3), 435-459.

- Khan, U., & Dhar, R. (2007). Where there is a way, is there a will? The effect of future choices on self-control. *Journal of Experimental Psychology: General*, 136(2), 277-288.
- Krasnovsky, T., & Lane, R. C. (1998). Shoplifting: A review of the literature. *Aggression and Violent Behavior*, 3(3), 219-235.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480-498.
- Lerner, M. J. (1980). "The belief in a just world," in *The Belief in a Just World: A Fundamental Delusion*, ed. Lerner M. J., editor. (New York, NY: Springer), 9–30.
- Levav, J., & McGraw, A. P. (2009). Emotional accounting: How feelings about money influence consumer choice. *Journal of Marketing Research*, 46(1), 66-80.
- Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421-55.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6), 633-644.
- Miller, D. T., & Effron, D. A. (2010). Chapter three-psychological license: When it is needed and how it functions. *Advances in experimental social psychology*, 43, 115-155.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of personality and social psychology*, 81(1), 33- 43.
- Moore, S. G., & Lafreniere, K. C. (2020). How online word-of-mouth impacts receivers. *Consumer Psychology Review*, 3(1), 34-59.
- Muncy, J. A., & Vitell, S. J. (1992). Consumer ethics: An investigation of the ethical beliefs of the final consumer. *Journal of business Research*, 24(4), 297-311.
- Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research*, 54(4), 572-588.
- Peer, E., Acquisti, A., & Shalvi, S. (2014). "I cheated, but only a little": Partial confessions to unethical behavior. *Journal of Personality and Social Psychology*, 106(2), 202.
- Pruckner, G., & Sausgruber, R. (2006). *Trust on the streets: A natural field experiment on newspaper purchasing* (No. 06-01). University of Copenhagen. Department of Economics.

- Ruedy, N. E., Moore, C., Gino, F., & Schweitzer, M. E. (2013). The cheater's high: The unexpected affective benefits of unethical behavior. *Journal of Personality and Social Psychology*, 105(4), 531-548.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychological science*, 20(4), 523-528.
- Schurr, A., Ritov, I., Kareev, Y., & Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgment and Decision Making*, 7(6), 679-688.
- Schwabe, M., Dose, D. B., & Walsh, G. (2018). Every Saint has a Past, and Every Sinner has a Future: Influences of Regulatory Focus on Consumers' Moral Self-Regulation. *Journal of Consumer Psychology*, 28(2), 234-252.
- Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, 25(2), 185-201.
- Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181-190.
- Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, 23(10), 1264-1270.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125-130.
- Sheeran, P. (2002). Intention—behavior relations: A conceptual and empirical review. *European review of social psychology*, 12(1), 1-36.
- Shu, L. L., & Gino, F. (2012). Sweeping dishonesty under the rug: how unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*, 102(6), 1164-1177.
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3), 330-349.

- Singh, P., & Twalo, T. (2015). Mismanaging unethical behaviour in the workplace. *Journal of Applied Business Research (JABR)*, 31(2), 515-530.
- Sniehotta, F. F., Scholz, U., & Schwarzer, R. (2005). Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. *Psychology & Health*, 20(2), 143-160.
- Tangney, J. P. (2003). Self-relevant emotions. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 384–400). New York: Guilford Press.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of personality and social psychology*, 78(5), 853-870.
- Tracy, J. L., & Robins, R. W. (2004). Putting the Self Into Self-Conscious Emotions: A Theoretical Model. *Psychological Inquiry*, 15(2), 103-125.
- Tracy, J. L., & Robins, R. W. (2007). The self in self-conscious emotions: A cognitive appraisal approach. In J.L. Tracy, R. W. Robins, & J.P. Tangney (Eds.) *The self-conscious emotions: Theory and research*, (pp. 3-20). New York: Guilford Press.
- Van Winden, F., & Ash, E. (2012). On the behavioral economics of crime. *Review of Law and Economics*, 8(1), 181-213.
- Wenzel, K., Schindler, S., & Reinhard, M. A. (2017). General belief in a just world is positively associated with dishonest behavior. *Frontiers in Psychology*, 8, 1770.
- Wilson, A. E., & Darke, P. R. (2012). The optimistic trust effect: Use of belief in a just world to cope with decision-generated threat. *Journal of Consumer Research*, 39(3), 615-628.
- Wirtz, J., & McColl-Kennedy, J.R. (2010). Opportunistic customer claiming during service recovery. *Journal of the Academy of Marketing Science*, 38(5), 654-675.
- Xu, H., Bègue, L., & Bushman, B. J. (2012). Too fatigued to care: Ego depletion, guilt, and prosocial behavior. *Journal of Experimental Social Psychology*, 48(5), 1183-1186.
- Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, 313(5792), 1451-1452.

תקציר

התנהגות לא-אתית היא תופעה רווחת, בעלת משמעות כלכלית נרחבת לארגונים ולחברה בכלל. עבירות כגון גניבה מחנויות, שימוש במשאבי חברה לצרכים אישיים, העלמת מס, והונאות ביטוח הן דוגמאות למעשים שאנשים רגילים מבצעים באופן קבוע המסתכמים בנזק כספי משמעותי. המחקר על התנהגות לא-אתית התמקד לרוב בסיבות להתנהגויות אלו וכיצד ניתן למנוע אותן. אולם, מחקרים רבים התעלמו במרבית המקרים מההשפעה האפשרית של התנהגות לא-אתית על השיפוט של קורבן המעשים הלא-אתיים. לדוגמה, אם בעל עסק עצמאי לא מדווח על כלל ההכנסות שלו, האם זה ישפיע על השיפוט שלו כלפי רשויות הממשלה והמדינה? אם יתבקש, האם יספק הערכה חיובית יותר או שלילית יותר של רשויות המדינה? תיאוריות נוכחיות מספקות ניבויים סותרים לשאלות אלו. מצד אחד, מחקרים הראו שאנשים חווים רגשות אשמה לאחר שהתנהגו בצורה לא-אתית, דבר שעלול להוביל לפעולות מתקנות והתנהגות פרו-חברתית. לפי גישה זו, אנשים שהתנהגו בצורה לא אתית עשויים לשפוט את הקורבן שלהם בצורה חיובית יותר כדי לפצות אותם על המעשה. מצד שני, כאשר אנשים מתקשים להצדיק את המעשים הפסולים שביצעו, הם עשויים להרחיק עצמם מהמעשה על ידי ביקורת חריפה על אחרים במטרה להציג את עצמם כמוסריים. לפיכך, גישה זו מנבאת שאנשים שהתנהגו בצורה לא אתית עשויים לשפוט את הקורבן שלהם בצורה שלילית יותר כדי לשקם את הדימוי העצמי המוסרי שלהם.

בעבודה זו, אני מעמתת בין שני ניבויים אלו כדי להבין טוב יותר את ההשפעה האפשרית של התנהגות לא אתית – כלומר, כיצד אנשים שופטים ומחווים דעה על הקורבנות של המעשים הלא-אתיים שלהם. כמו כן, אני בוחנת האם ובאיזו צורה שיפוט הקורבן של מעשה הרמייה מושפע מהיכולת להצדיק את ההתנהגות הלא אתית. לבסוף, אני בוחנת האם התנהגות לא-אתית משפיעה על הרגשות המובעים בביקורות הכתובות ועל נכונות הקוראים להיות בקשר עם הקורבן של המעשים הלא אתיים. התופעה נבחנה במספר סביבות פעילות, הן במקרים אמיתיים והן באופן מקוון. הממצאים מצביעים על כך שאנשים שמתנהגים בצורה לא-אתית שופטים את הקורבן שלהם בצורה שלילית יותר, ככל הנראה כדי להרחיק את עצמם מהמעשה ובכך להצדיק את ההתנהגות שלהם. התוצאות היו עקביות ונמצאו בקרב עובדים ששפטו את המעסיקים שלהם בצורה פחות חיובית לאחר שנזכרו בהתנהגות הלא אתית שלהם במקום העבודה, ובקרב עצמאיים ששפטו רשויות מס בצורה שלילית יותר לאחר שנזכרו בדוחות המס השקריים שלהם. באופן דומה, בהקשר שיווקי, משתתפים שבחנו אפליקציה חדשה, שפטו אותה באופן שלילי יותר ככל שהם רימו יותר. בנוסף, במסגרת המחקר מצאתי שהתנהגות לא-אתית לא משפיעה על שיפוט הקורבן כאשר הרמייה היא ברורה וקשה להצדקה. ממצא זה תומך במנגנון הריחוק (Distancing) כמנגנון המניע את האפקט. לבסוף, מצאתי שכאשר אנשים מתנהגים בצורה לא מוסרית, הם נוטים לבטא תחושות שליליות יותר בביקורות הכתובות שלהם ובכך מפחיתים את נכונות הקוראים להיות בקשר עם קורבן ההתנהגות הלא אתית. עם זאת, ביקורות שנכתבו על ידי משתתפים שקיבלו הזדמנות לרמות, אך עמדו בפיתוי, העלו את נכונות הקוראים להשתמש במוצר בהשוואה לביקורות שנכתבו על ידי משתתפים שלא יכלו לרמות.

לעבודה זו תרומה תיאורטית משמעותית מכיוון שהיא מדגישה היבט שלא נחקר עד כה בהתנהגות לא-אתית. המחקר בוחן את ההשפעה שיש להתנהגות לא אתית על השיפוט וההערכה של קורבן המעשים ולא בסיבות להתנהגות זו ולדרכים לצמצם אותה. במובן זה, המחקר מחשיב התנהגות לא-אתית כגורם ולא כתוצאה. בנוסף, המחקר תורם לתחום ההולך וגדל של תקשורת מפה לאוזן (word of mouth) על ידי חשיפת גורם חדש שיכול להטות באופן שיטתי דירוגים וחוות דעת. מכיוון שהתנהגות לא אתית היא נפוצה, סביר

להניח שביקורות מקוונות רבות נכתבות על ידי אנשים שהתנהגו בצורה לא אתית. ממצאי מחקר זה מרמזים על כך שביקורות אלו עשויות להיות מוטות בשל הנטייה של אנשים המתנהגים באופן לא מוסרי לשפוט בחומרה את קורבן הרמייה שלהם.

בנוסף, לעבודה זו יש השלכות חשובות על ארגונים, קובעי מדיניות ומשווקים בגלל ההשפעות השליליות המתמשכות שיש להתנהגות לא אתית. מאחר שארגונים כפופים לדעת הקהל, שיפוט שלילי של אנשים לא מוסריים עשוי להשפיע על התדמית שלהם בעיני הציבור, על התוצאות העסקיות של החברה, וכתוצאה מכך על יכולתם לגייס עובדים טובים, לספק שירותים איכותיים או לשמור על האינטרסים של בעלי העניין. בנוסף, הממצאים מדגישים את החשיבות של מדיניות שמטרתה להרתיע אנשים מלהתנהג בצורה לא מוסרית, מכיוון שנקיטה במדיניות מקלה עשויה לחשוף חברות לביקורות שליליות יותר ברשת. לבסוף, המחקר מציע השלכות הנוגעות להגנת הצרכן ולמדיניות סחר הוגן. מכיוון שלהתנהגות לא אתית יש השלכות משמעותיות במובן של תקשורת "פה-לאוזן", יש ליידע את האנשים לגבי הטיות אפשריות בתהליך קבלת ההחלטות שלהם בזמן שהם שוקלים קניית מוצר או מתקשרים עם ארגונים. לבסוף, על קובעי המדיניות לשקול מדיניות שמטרתה להגן על האינטרסים של הארגונים בשל התוצאות השליליות של התנהגויות לא אתיות.

עבודה זו נעשתה בהדרכתו של פרופסור איל פאר

כיצד אנשים שמתנהגים בצורה לא אתית שופטים את קורבנות מעשיהם?

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת

נורית הוד

הוגש לסנט האוניברסיטה העברית בירושלים

ספטמבר/2021