



Do Moral Reminders Shape Ethical Behaviour in the Workplace?

Thekla Maria Schmidt (s2586479)

University of Groningen

Date: 26-06-2018

Master Thesis

Work, Organizational and Personnel Psychology

Faculty of Behavioural and Social Sciences

Supervisor: Dr. E. F. Rietzschel

Second Evaluator: Dr. A. C. Keller

External Supervisor: Dr. A. Pittarello – Brooklyn College, CUNY, New York (USA)

Acknowledgements

Firstly, I would like to express my sincere gratitude to Andrea Pittarello for his persistent support, expertise, and advice in all stages of conducting my research throughout the past academic year. I also would like to extend my greatest appreciation to Ruth Mayo, who shared her knowledge in stimulating discussions, provided another perspective, and offered helpful suggestions on this research. I owe my gratitude to Eric Rietzschel for his guidance and insightful comments on my research. Also, I am grateful for the pleasant collaboration with the company that provided me the opportunity to conduct this research and get access to the sample for my research. Last but not least, I am immensely grateful for the continuous encouragement and help that I got from my friends and family.

Abstract

Dishonesty in the workplace is a major concern faced by modern businesses. The present study suggests possibilities to tackle this problem by varying the wording of moral reminders and providing information about the people that are harmed by dishonesty. I tested the effect of written moral reminders worded in either an affirmative (“Please pay here if you took a snack”) or negative (“Please do not forget to pay if you took a snack”) way among 103 German offices that are regularly supplied with snack boxes and where employees purchase snacks by means of an honour system. Additionally, an image of the people harmed by dishonesty was attached to the snack boxes, depicting either a single or a group of individuals. The two types of messages and images were presented in combination as well as separately, resulting in eight between-subject conditions. For each office I computed a difference score by subtracting the payment percentage of the baseline level (prior to the intervention) from the one following the intervention. Results showed that payment rates tended to increase following the affirmative vs. negative message, and when the image depicted a single vs. group of individuals. This study hints toward important practical and cost-effective implications for businesses to curb dishonesty in the workplace.

Keywords: ethics, moral reminders, negotiations, identifiable victim

Table of Contents

Introduction.....	5
Dishonesty Prevention Strategies.....	7
Written moral reminders.....	7
Images of people harmed by dishonesty.....	12
Method.....	15
Research Design and Participating Offices.....	15
Dependent variables.....	15
Participating offices.....	16
Materials.....	16
Procedure.....	17
Results.....	17
Discussion.....	20
Theoretical Contributions.....	21
Strengths, Limitations, and Future Research.....	24
Practical Implications.....	26
Conclusion.....	27
Footnotes.....	28
References.....	29
Tables.....	34
Figures.....	35
Appendix A.....	36
Appendix B.....	38

Do Moral Reminders Shape Ethical Behaviour in the Workplace?

Imagine an ordinary day at your office. Next to your desk there is a cupboard fully stocked with supplies that are meant for office purposes only. However, you often notice that some of your colleagues regularly take from this cupboard for private purposes. One day someone takes home a pad of post-it notes, another day someone takes home a highlighter, and on the next day another colleague is using a company envelope to mail a personal letter. Unfortunately, such scenario is a pretty accurate representation of reality. Indeed, research showed that employee theft is one of the most widespread, prevalent, and costly concerns faced by today's businesses (Hollinger & Davis, 2006). Approximately 75% of employees have reportedly stolen from their employers at least on one occasion and many of these do so regularly (Jones, 2012). Annual financial losses to businesses resulting from employee theft have been estimated to be as high as \$20 to \$40 billion (Cragle, 2015). Of all business failures, 30% can be attributed to employee theft (Kuratko, Hornsby, Naffziger, & Hodgetts, 2000). People tend to rationalize petty theft by believing that no one is harmed by such actions (Anand, Ashforth, & Joshi, 2004). However, accumulated over time, even theft of small items can cause massive financial damages (Mazar, Amir, & Ariely, 2008). Without a doubt, these numbers are alarming and call for urgent solutions on what can be done to curb unethical behaviour in the workplace. Here, I suggest possible ways to tackle this problem: Varying the wording of moral reminders and providing information about the people that are harmed by the unethical conduct.

In the past, many companies have invested in expensive ethics programs to fight unethical behaviour. However, these programs are often short-lived, and rarely effective (Greenberg & Barling, 1996; Richards, 1999; Tenbrunsel & Messick, 2004). One explanation for such failure is that these programs are targeted toward stricter control structures and largely neglect the psycho-

logical processes underlying the unethical behaviour. Bazerman and Tenbrunsel (2011) argue that – when behaving unethically pays – people fail to pay sufficient attention to their misbehaviour. This occurs for a variety of reasons: The excessive focus on the bottom-line of the company, the presence of temptation, and ambiguous rules of conduct sway people away from behaving ethically and instead lead them to behave in a self-serving manner while maintaining a positive self-concept (Mazar et al., 2008; Pittarello, Leib, Gordon-Hecker, & Shalvi, 2015). One approach to direct people's attention to their (mis)behaviour is by written messages, because they can activate social norms and thereby remind people of the norm violation (Cialdini, 2003). Another approach that might increase people's attention to their (mis)behaviour is the presence of images of the people harmed by the unethical conduct, which – in the field of prosocial behaviour (e.g., Kogut & Ritov, 2005) – has been found to increase the concerns for those in need.

The present study examines whether the effectiveness of such measures in limiting people's unethical behaviour applies in the field in a context where employees purchase snacks by means of an honour system. This is a sales method by which consumers purchase a product on a self-service basis and pay for the product without being actively monitored (Brudermann, Bartel, Fenzl, & Seebauer, 2015). Hence, the system relies on the honesty of consumers to pay the required amount of money. Honour systems are commonly used for the sale of coffee or snacks in company canteens. If written moral reminders and images depicting the people harmed by dishonesty will indeed reduce money embezzlement, these strategies may help conveying desirable standards of behaviour and promote an ethical climate. Ultimately, they can successfully reduce financial losses caused by employee theft in a simple and cost effective manner.

Dishonesty Prevention Strategies

In the past, one approach that scientists have applied to design interventions aimed at reducing people's dishonesty was to use moral reminders. Moral reminders are means of directing people's attention towards moral standards. The attention to one's moral standards confronts people with the consequences of their behaviours, oftentimes resulting in more honesty (Mazar et al., 2008). Some of the most relevant field studies using moral reminders investigated the effect of artificial surveillance cues on people's unethical behaviour. For instance, several studies tested whether an image depicting eyes on the wall leads individuals to behave as if they are being observed by a real person by increasing their concerns for their reputation (e.g., Bateson, Callow, Holmes, Roche, & Nettle, 2013; Bateson et al., 2015). However, Cai and colleagues (2015) found that, compared to a neutral image, the image of observant eyes did not increase honesty, because – at least in anonymous settings like in this case – implicit reputation cues are not enough to increase people's attention to their wrongdoing. Instead, explicit moral reminders, like written messages, might be necessary in tackling dishonest behaviours.

Written moral reminders. Written messages reminding people of behaving honestly can be formulated in different ways. For instance, in a study that investigated the prevention of bicycle theft, an image of observant eyes was complemented by a related written message telling the thieves that they are being watched (i.e., "Cycle thieves, we are watching you") (Nettle, Nott, & Bateson, 2012). In this setting, theft was effectively reduced compared to locations without any moral reminders. To study the effect of moral reminders in more detail, several studies investigated written messages alone, showing that – to an extent – written moral reminders can effectively curb dishonesty. For example, consumers paid more for newspapers in an honour system when being exposed to a written message that was directly appealing to their morale

(e.g., “Thank you for being honest”) (Pruckner & Sausgruber, 2013). Therefore, the use of written moral appeals has been found to be an effective dishonesty prevention strategy.

Furthermore, as mentioned above, social norms can considerably influence ethical behaviours. The general term *norm* refers to two distinct forms of information – *descriptive* and *injunctive* information. Descriptive norms are standards outlining which behaviours are commonly performed, whereas injunctive norms are standards outlining which behaviours are commonly (dis)approved (Cialdini, 2003). Research on descriptive norms is more widely studied, showing that the use of these norms can amplify people’s conduct when the desired behaviour is frequently observed in the population. For instance, towel reuse in US hotels increased when descriptive norms (e.g., “The majority of guests reuse their towels”) were used compared to traditional appeals, which mainly addressed environmental protection (Goldstein, Cialdini, & Griskevicius, 2008). However, the positive effect of descriptive norms on people’s behaviour is limited to individuals who already adopt the desired behaviour and can even be reversed for people that are adopting the behaviour to a lower extent (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007).

Moreover, in settings where undesirable behaviours are more frequently observed, the use of injunctive norms (e.g., “Please don’t remove the petrified wood from the park”) seems to be more effective than descriptive norms (e.g., “Many past visitors have removed the petrified wood from the park, changing the state of the Petrified Forest”) (Cialdini et al., 2006). In these settings, descriptive norms highlight the undesirable behaviour and therefore might even further enhance this conduct. Keizer and colleagues (2008; 2011) propose that the effectiveness of injunctive norms is also limited. According to their research, individuals tend to violate injunctive norms when others’ less socially desirable behaviour is directly salient. Specifically, people

broke the rules (in this case, spraying graffiti) when there were signs of other people doing so (descriptive norm) even in the presence of prohibition signs (injunctive norm, in this case an “anti-graffiti” sign). Therefore, the incongruence between the descriptive and injunctive norm led to a violation of the injunctive norm. However, this does not imply that descriptive norms are generally more powerful than injunctive norms. Rather, in case of contradictions between an injunctive and descriptive norm, the attentional focus on one or the other form of normative information determines with which norm individuals comply (Cialdini, Reno, & Kallgren, 1990).

In the present study, the written messages target employees’ non-accurate payments of snacks, meaning that they either do not pay for what they purchased, or, paid a lower amount of money than the actual price. This dishonest behaviour occurs to an undesirable extent. However, payments are inserted into a wooden cashbox, making it more difficult for employees to ascertain whether others accurately paid for the snacks. Therefore, in this study, written moral reminders are formulated as injunctive norms, following Cialdini and colleagues’ (2006) suggestion.

Although these norms can be formulated in many different ways, one question that remains open, and is largely neglected in the current literature, is: What is the most effective wording to craft such reminders? Could it be that, despite our best intention, using a particular wording might actually backfire and increase the undesirable behaviour? This is my question of interest.

Messages can be expressed in an affirmative or negative way. Affirmatively worded injunctive norms highlight a desirable behaviour (e.g., “Please throw your litter in the appropriate bins”), whereas negatively worded injunctive norms highlight an unwanted behaviour (e.g., “Please do not throw your litter on the streets”) (Shannon, 2000). One intriguing insight from social cognitive literature suggests that, despite intending to convey the same meaning,

such subtle changes in the wording of messages can activate different concepts. For instance, messages expressed as negations might activate the opposite of the intended meaning. One study investigated the possibility of negative messages to activate false memories. More specifically, individuals listened to an audio clip describing a house. When they were presented with a negative statement about an object (e.g., “There is no porch”) and questioned about the presence of the object one week later, they recalled the nonexistent object as in fact existing (Maciuszek & Polczyk, 2017). Hence, the negation activated a false representation of the object after some time. A different study examined in what way the cognitive processing of negations differs from that of affirmations. According to that study, negative statements (e.g., “This lawyer is not a shark”) are initially represented as affirmations (e.g., “This lawyer is a shark”), but in later stages of comprehension the negation is integrated (Hasson & Glucksberg, 2005). This indicates that negative messages activate the negated concept, and thereby foster associations contrary to the message’s intended meaning, before being completely and correctly processed.

Mayo and colleagues (2004) investigated the processing of negative messages in more detail. They refer to two models that explain the activation of different concepts in negations and that apply depending on the type of negated information. The *schema-plus-tag model* states that negative messages (e.g., “Jim is not responsible”) are processed in two steps. First, individuals process the core supposition (schema; e.g., “responsibility”) of the message. Subsequently, this supposition is completed with the negation tag (“not”). However, the negation tag could be disregarded, and thus the message facilitates incorrect associations. In contrast, affirmative messages (e.g., “Jim is responsible”) activate associations congruent with the message’s intended meaning. This model applies in cases where no clear opposite of the negated schema exists (e.g., responsible or creative). In contrast, according to the *fusion model*, the schema and negation tag

are merged and the message is thereby transformed into an affirmation, reinforcing congruent associations. For instance, the statement “Person A is not guilty” might be processed as “Person A is innocent”. This model applies when a clear opposite of the negated schema exists (e.g., warm vs. cold).

I suggest that, when it comes to (un)ethical behaviour, the activation of associations incongruent with the intended meaning can have dramatic consequences. In other words, negating an unwanted behaviour might activate concepts related to dishonesty, and thus increase it. This can particularly be the case when messages offer potential justifications for the dishonesty. Indeed, research found that when self-interest is pitted against being honest, people bend the rules to the extent to which they can justify or reframe their behaviour in acceptable manners (Barkan, Ayal, & Ariely, 2015; Pittarello et al., 2015; Schweitzer & Hsee, 2002; Shalvi, Dana, Handgraaf, & De Dreu, 2011).

The present study assesses the effects of the negative message “Please do not forget to pay if you took a snack” vs. the affirmative message “Please pay here if you took a snack”. However, although the negated verb “forget” semantically possesses the clear opposite “remember”, I expect this message to be processed according to the schema-plus-tag model. I draw this conclusion from research in the field of social psychology, behavioural ethics, and management, which proposes that, when executing unethical acts, people tend to engage in self-deception (Tenbrunsel & Messick, 2004). By disguising the moral implications of their conduct, they maintain a coherent moral self-image. Additionally, people tend to search for justifications for their unethical conduct (Tenbrunsel, 1995). Hence, individuals might be more dishonest when being exposed to the negative reminder in comparison to the affirmative reminder because the negative reminder provides a possible justification for their inaccurate payments (because they

simply forgot to pay), activating the schema of forgetting. This reasoning is in line with research on *motivated forgetting*, which suggests that people suppress the moral rules that would otherwise limit their unethical conduct (Kouchaki & Gino, 2015). In contrast, the affirmative message in the present research simply reminds individuals to act according to the moral rules, challenging the self-deceptive mechanisms by making it more difficult for individuals to suppress the rules. Taking these considerations into account lead to my first hypothesis:

H1: *Affirmative messages lead to higher honesty levels than negative messages.*

Images of people harmed by dishonesty. In addition to the possibility of reducing dishonest behaviours through written reminders, images might serve as another effective dishonesty prevention strategy. Images convey affect and meaning. This often makes them more powerful than simply conveying written messages, which are processed more rationally (Slovic, 2010). In the present study, images depicting the people harmed by dishonesty are expected to deter people from behaving dishonestly. This assumption is drawn from research on the effect of victim anonymity on unethical behaviour, showing that individuals were more reluctant to conduct ethically questionable behaviours, like withholding life-saving products, allocating payments unfairly, or cheating overtly, when the victim of the wrongdoing was identified by a name compared to an anonymous victim (Yam & Reynolds, 2016). This phenomenon relates to the *identifiable victim effect*, which states that an identifiable victim engenders more empathy in the perceiver than an anonymous victim and thereby enhances the perceiver's willingness to exhibit behaviours that will lessen the victim's detriment (Jenni & Loewenstein, 1997). Similarly, another study demonstrated that individuals judged others' behaviours as more unethical when the person that was harmed was identifiable by a name rather than anonymous (Gino, Shu, & Bazerman, 2010). Since the presence of a picture of an individual reduces anonymity, it may be

expected that images of the individuals who are affected by the dishonesty, similar to other identifying information, will also prevent unethical conduct in the people exposed to them. In the context of this study, the people that are harmed by not paying in full for the snacks are the employees of the company that is delivering the snack boxes. To the best of my knowledge, this is the first study testing the identifiable victim effect in the context of unethical behaviours in the workplace.

Furthermore, previous research on prosocial behaviours has considered the effect of the number of victims depicted. According to this research, the number of victims depicted on a photo elicits different psychological responses in the perceiver. For instance, research on people's willingness to donate indicates that an image depicting a single individual evokes a stronger psychological response than an image depicting a group of individuals (Kogut & Ritov, 2005), because single individuals are perceived as more coherent entities (Hamilton & Sherman, 1996; Susskind, Maurer, Thakkar, Hamilton, & Sherman, 1999) and trigger stronger affective reactions than groups of individuals (Slovic, 2010). Therefore, an image of a single individual might elicit a stronger sense of morale than an image of a group of individuals. Consequently, people might cheat a single individual less than a group of individuals when these individuals will suffer the effect of dishonesty. This study will advance our knowledge in that it examines whether the effects found in contexts aiming to activate prosocial behaviour are also applicable to contexts aiming to curb unethical behaviour.

The disparity between the perception of single individuals and that of groups is likely to become larger the more detailed the available information is. Accordingly, when images of single individuals vs. groups are accompanied by other pieces of identifying information (e.g.,

name, age) they increase feelings of empathy and distress (Kogut & Ritov, 2005). This led me to predict that:

H2: Images depicting single identified individuals lead to higher honesty levels than images depicting groups of identified individuals.

If messages framed as affirmations increase honesty more compared to those framed as negations and if people feel stronger reactions when presented with an image of a single individual vs. a group and thereby behave more honestly, then it is reasonable to assume that some written messages will be even more effective when they are linked to specific images. In other words, the positive effect of an affirmative message on people's honesty might be strengthened when this message is paired with an image of a single person harmed and should therefore be the most effective strategy to curb dishonesty. Additionally, the negative effect of a negative message on people's honesty might be strengthened when this message is paired with an image of a group of people harmed and should therefore be the least effective strategy to curb dishonesty. Thus, I predict an interaction effect between the wording of the message and the number of individuals depicted on an image. Formally put: The effect of the wording of moral reminders on honesty is amplified by the number of people that are harmed and depicted on an image, such that

H3a: The effect of affirmative messages on honesty is amplified when paired with images of single individuals.

H3b: The effect of negative messages on honesty is amplified when paired with images of groups of individuals.

To test the proposed hypotheses, I conducted a field study in which I recruited a company that delivers snack boxes to offices in Germany using an honour system. The effectiveness of

written moral reminders and images depicting the people harmed by dishonesty in reducing employees' unethical conduct will be represented in an increase in payment rates.

Method

Research Design and Participating Offices

The present study employed a 2 (affirmative vs. negative message) x 2 (single vs. group of individuals) between-subjects design which results in four groups reflecting all possible combinations. In addition, four control conditions were included in which either only a message or only an image was displayed (1) affirmative message, 2) negative message, 3) image of a single individual, 4) image of a group of individuals). The resulting eight groups (4 treatment conditions plus 4 control conditions) were randomly assigned to the 903 most lucrative client offices of the delivery company that are regularly supplied with snack boxes. Therefore, this study represents a field experiment with two manipulated independent variables (*Type of Message* and *Type of Picture*) and random assignment to the groups. The last payment prior to the implementation of the moral reminders was assessed as a baseline condition.

Dependent variables. Before conducting the analyses, I computed two variables of interest as my dependent variables. The first, called *Percentage Paid*, corresponds to the proportion of the total amount of money deposited in the cashbox at the end of the exchange period relative to the amount expected assuming total honesty (based on the prices of the consumed snacks) during the intervention. The same percentage was computed for the last exchange period prior to the intervention, which corresponds to the *Percentage Paid Baseline*. To test the effect of my intervention on payment rates, I subsequently computed the difference between the *Percentage Paid* and the *Percentage Paid Baseline*, constituting the second dependent variable called *Difference Percentage Paid*. A positive score of this variable would indicate an increase in

payment rates from baseline to the intervention, whereas a negative score would indicate that payment rates dropped during the intervention.

Participating offices. Since the length of the snack boxes' exchange periods differs between offices, the money was collected after two to eight weeks. Additionally, only a certain amount of offices could be supplied with snack boxes each day depending on the timely resources of the delivery company. Consequently, the date of the initial implementation of the messages and images as well as the date of the money collection varied between offices. The period of the present study lasted 12 weeks, from February 2018 until May 2018. Due to these constraints, measurements of only a subset of 103 out of the 903 offices could be collected at this point in time. Hence, only a subgroup of offices in which the intervention was implemented could be analysed.

The 103 offices are active in different sectors, from legal services, consulting, education, media, healthcare, consumer products, tourism, financial services, IT and telecommunication, to construction, industrial goods and services, logistics, automotive industry, and energy supply. The number of employees having access to the snack boxes was not provided by 15 offices. From the remaining 88 offices, the mean number of employees is 42.41 ($SD = 32.00$), with a range from 6 to 150 employees. The total value of contents that were initially placed in the snack boxes differed between offices from € 112.00 to € 240.00, with a mean of € 135.90 ($SD = 29.47$).

Materials

The signs conveying the written moral reminders and images (see Appendix A) to the employees were attached to each cashbox. The written reminders were glued onto the cashboxes, conveying either the affirmative message "Please pay here if you took a snack", or the negative message "Please do not forget to pay if you took a snack". The signs with images depicting the

people affected by dishonesty were attached behind the cashboxes. These signs were consistently complemented with the words “Thank you from the XX-team”¹ and with a short message including the names and the tenure of the individuals depicted on the images. In the group condition, three employees of the delivery company were depicted.² In the single condition, only one of the three individuals who were depicted on the group photo was present. Importantly, the single vs. group pictures were identical, but only digitally modified to include one vs. three persons. An example of a snack box conveying a written moral reminder combined with an image is given in Appendix B.

Procedure

The study received ethical approval from the University of Groningen. The 903 most lucrative client offices of the delivery company were selected and randomly assigned to the eight groups. Subsequently, the snack boxes, with their particular moral reminder, were delivered to the offices. A sign attached to every snack box labeled the price of the snacks (mostly € .80, with some exceptions costing € .40). Employees are expected to insert the respective amount of money for their taken snacks into a cashbox included in the snack box. However, their payment is not directly monitored. Hence, consumers can take a snack and correctly pay for it (or too much), or they can take a snack without paying the correct amount of money (i.e., too less or no payment). When the money is collected, the boxes are getting refilled. However, for the present study, out of time constraints, payments from only one exchange period could be collected.

Results

Of the 103 offices, one was removed from the dataset because it indicated an implausibly high payment rate of 244%, which is 6.88 standard deviations above the average payment rate

($M = .79$, $SD = .24$) and thus greatly exceeds the common rule of thumb for outlier detection (i.e., greater than 3 standard deviations above the mean; Leong & Austin, 2006).

The statistical analysis was conducted with IBM SPSS Statistics version 25. First, I conducted a 2 x 2 analysis of covariance (ANCOVA) predicting the *Percentage Paid* during the intervention to analyse whether there are significant differences in payment rates between the implemented messages and images while controlling for the *Percentage Paid Baseline*. Levene's test was carried out and the assumption of homogeneity of variance was met. The Shapiro-Wilk test showed that the assumption of normality was not met. However, since none of the recommended transformations satisfactorily addressed this issue and since ANCOVA is a robust analysis to violations of normality (Levy, 1980), I used the non-transformed actual payment rates for this ANCOVA. Table 1 displays the means and standard deviations of all groups. Results revealed no significant main effect of the *Type of Message*, $F(2, 93) = 2.33$, $p = .10$, nor of the *Type of Picture*, $F(2, 93) = 1.25$, $p = .29$. Finally, the interaction between *Type of Message* and *Type of Picture* was not significant, $F(3, 93) = .25$, $p = .86$.

Next, I conducted a 2 x 2 analysis of variance (ANOVA) predicting the *Difference Percentage Paid*, that is, the pure effect of my intervention. This measure is more fine-grained because it determines the strength of the intervention by analysing whether the implemented messages and images evoked a change in payment rates compared to the last period prior to the intervention. Levene's test was carried out and the assumption of homogeneity of variance was met. The Shapiro-Wilk test showed that the assumption of normality was not met. However, since none of the recommended transformations satisfactorily addressed this issue and since ANOVA is a robust analysis to violations of normality (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010), I used the non-transformed differences in payment rates for this ANOVA.

Results showed a marginally significant main effect of the *Type of Message*, $F(2, 94) = 2.56$, $p = .08$, $\eta_p^2 = .05$. Contrasts analysis showed that the average difference in payment rates from baseline to intervention in the groups presented with the affirmative message ($M = .06$, $SD = .17$) differed marginal significantly from that in the groups presented with the negative message ($M = -.02$, $SD = .17$), in that the affirmative message tended to elicit an increase in payment rates and the negative message tended to elicit a decrease in payment rates from the baseline to the intervention ($M_{diff} = .09$, 95% CI $[-.01, .18]$, $p = .08$). This marginally supports my hypothesis that affirmative messages lead to higher honesty levels than negative messages (H1). The average difference in payment rates in the groups that were not presented with any message (i.e., only a picture) ($M = -.01$, $SD = .13$) did not significantly differ from that of the affirmative ($M_{diff} = -.07$, 95% CI $[-.19, .04]$, $p = .36$) and negative message ($M_{diff} = .02$, 95% CI $[-.09, .12]$, $p = 1.00$).

The main effect of the *Type of Picture* was significant, $F(2, 94) = 3.72$, $p = .03$, $\eta_p^2 = .07$. Here, contrasts analysis indicated that the average difference in payment rates in the groups presented with the image of a single individual ($M = .05$, $SD = .10$) significantly differed from that in the groups presented with the image of a group of individuals ($M = -.06$, $SD = .21$), in that the image of a single individual elicited an increase in payment rates and the image of a group elicited a decrease in payment rates from the baseline to the intervention ($M_{diff} = .11$, 95% CI $[-.01, .22]$, $p = .03$). This confirms my hypothesis that images depicting single identified individuals lead to higher honesty levels than images depicting groups of identified individuals (H2). The average difference in payment rates in the groups that were not presented with an image (i.e., only a message) ($M = .03$, $SD = .15$) did not significantly differ from that of the image of a single individual ($M_{diff} = -.03$, 95% CI $[-.12, .07]$, $p = 1.00$) and differed marginal significantly

from that of the image of a group ($M_{diff} = .09$, 95% CI [-.01, .18], $p = .08$). Figure 1 presents the averaged differences in payment rates between the baseline and intervention period across all groups.

The interaction effect between the *Type of Message* and the *Type of Picture* was not significant, $F(3, 94) = .14$, $p = .94$. Therefore, the hypotheses that regard the interaction between the wording of the message and the number of individuals depicted on an image (H3a and H3b) were not supported. However, although no significant interaction occurred, the findings show interesting trends in the predicted direction, with the affirmative message together with the image of a single individual showing the greatest increase in payment rates ($M = .10$, $SD = .10$) and the negative message together with the image of a group of individuals showing the greatest decrease in payment rates ($M = -.11$, $SD = .30$). Taken together, these results seem to suggest that affirmative messages and images of single individuals might be effective ways of curbing dishonesty.

Discussion

In this study, I tested whether written moral reminders conveying social norms and images depicting the people suffering from the unethical behaviour are effective strategies to curb dishonesty. Specifically, I examined how the wording of a written moral reminder and the number of people that are harmed and depicted on an image influence people's (dis)honesty – here, the tendency to avoid paying for snacks at work. Results suggest that payment rates tended to increase following an affirmative message telling employees to pay if they took a snack, meaning that activating a desirable behaviour might increase honesty. Conversely, payment rates tended to decrease following a negative message telling employees not to forget to pay if they took a snack, meaning that negating an undesirable behaviour might paradoxically increase

dishonesty. While these trends are pointing towards the predicted direction, the marginal effect of the *Type of Message* calls for cautious conclusions and only marginally supports my prediction that affirmative messages lead to higher honesty than negative messages (H1).

Additionally, payment rates increased in the presence of an image of a single vs. group of individuals complemented by the name and tenure of the person(s) depicted, confirming my prediction that images depicting single individuals accompanied by additional information lead to greater honesty than images depicting groups of individuals accompanied by additional information (H2). Surprisingly, the image of a group even increased employees' dishonesty. No interaction emerged between the wording of the message and the number of individuals depicted on the image, failing to support H3a and H3b.

Common wisdom might suggest that asking people to refrain from undesirable behaviours should be the most effective way to limit such behaviours. The data reported here appear to suggest that doing so might actually backfire, at least when the wording of the message provides potential justifications for the misconduct. On the other hand, highlighting a desirable behaviour seems to lead people towards desirable (here: moral) behaviours. Thus, while there are several ways to convey the same meaning, choosing the right wording is far from trivial. Even more evidently, this study showed that displaying more than one person harmed by dishonesty seems to actually increase dishonesty, while a single person promotes more honesty. Based on these findings, I suggest that it is best to keep reminders both personal and affirmative.

Theoretical Contributions

This study provides important theoretical contributions to research in the field of behavioural ethics. Specifically, I focused on ethical behaviour in the workplace and investigated strategies to curb employees' dishonesty in a real life working environment. One strategy that is

aiming to reduce dishonesty in the workplace is the use of written moral reminders. So far, research on written moral reminders has not paid sufficient attention to the exact wording of such messages. The present study uses a multidisciplinary approach to fill this gap by integrating two different, yet related fields: Behavioural ethics and social cognition. In line with social cognitive findings, which indicate that negations might trigger associations that are contrary to the message's intended meaning (Mayo et al., 2004), the negative reminder used in my study tended to increase employees' dishonest behaviours, meaning that it elicited a behaviour that is contrary to the intended meaning.

The present study provided unique insights by pointing towards a potential boundary condition of the fusion model that Mayo and colleagues (2004) discuss, stating that negative messages activate associations congruent with the message's intended meaning when the negated schema possesses a clear opposite (e.g., warm vs. cold). In my study, the negative message tended to increase dishonesty, suggesting that the message might have activated associations incongruent to the message's intended meaning, even though the term "forget" semantically possesses an opposite (i.e., remember). This might suggest that negated schemata that are possessing a clear opposite might not always lead to associations congruent with the message's intended meaning. Here, I cautiously propose that the activation of a justification might have led people to process the message according to the schema-plus-tag model, resulting in more dishonesty. More specifically, the negative reminder might have increased people's dishonesty by activating the justification of "forgetting" for people's failure to pay (see Barkan et al., 2015; Kouchaki & Gino, 2015) and thereby disregarding the negation tag ("not"). In other words, the activation of a justification might have served as the underlying mechanism that led employees to neglect the negation, in turn decreasing honesty.

Furthermore, the present study utilized the ability of images to influence people exposed to them, thus integrating previous research on prosocial behaviour to the growing field of behavioural ethics. In line with work by Kogut and Ritov (2005) showing that a single person in need increases psychological responses in the perceiver, I cautiously argue that the same processes evoked an increase in employees' honesty. Surprisingly, in this study, the image of a group of people that are harmed increased employees' dishonesty. Therefore, the contextual difference of activating prosocial behaviour compared to reducing unethical behaviour might constitute a crucial factor. More specifically, in a context where people are asked for donations, images of single individuals effectively rose donations while images of groups did not elicit any change in donations compared to no image (Kogut & Ritov, 2005). In contrast, in contexts that address people's unethical behaviour, the positive influence of an image of a single individual on people's honesty might not be the only effect that can be observed. Instead, the image of a group of individuals led to an increase in dishonesty.

One possible explanation for the increase in dishonesty when being presented with an image of a group might be that people believed that the harm is distributed among several group members and thereby not as detrimental as the harm that is inflicted on a single person suffering from the effect of dishonesty. By believing that the damage caused by dishonesty is minor to the group compared to a single person, employees might have used a rationalization technique referred to as *denial of injury* (Anand et al., 2004). Hence, this study offers an interesting hint to a potential underlying mechanism that might lead people to behave less honestly when being presented with an image of a group vs. a single individual harmed by dishonesty.

Moreover, the combination of written moral reminders and images depicting the people harmed by dishonesty is a novel contribution to research. While the interaction between the wor-

ding of written reminders and the number of people depicted on an image was not supported, the pattern of results seems to be congruent with my prediction, showing that the affirmative reminder together with the image of a single person increased payment rates the most and the negative reminder together with the image of a group of people decreased payment rates the most.

Strengths, Limitations, and Future Research

This study has both advantages and disadvantages. Importantly, the use of a field study represents a major advantage. Participants' behaviour was observed and recorded in their usual working environment, which means that this study possesses high external validity compared to a laboratory study (Kaplan & Saccuzzo, 2013). However, due to a pre-selection of the 903 most lucrative client offices of the delivery company, the question remains open as to whether the results hold true if the messages and images would be implemented in all client offices. Since the length of exchange periods of the less lucrative clients exceeded the time frame of this study, this pre-selection could not be avoided. Moreover, due to time constraints on the part of the delivery company, not all delivered snack boxes could be collected at this point in time. Consequently, only a subset of 103 from a total of 903 offices in which the intervention was implemented, could be used. While the trends of the effect of the *Type of Message* on honesty are in line with my prediction, the marginal effect calls for a need to replicate the findings with a larger sample. Moreover, future research could examine whether the effectiveness of the moral reminders changes over time by computing a series of exchange periods across a longer time frame. A repeated observation might shed light on the robustness of the effects demonstrated in this study and might potentially yield the predicted interaction between the wording of a moral reminder and the number of people that are harmed and depicted on an image.

Furthermore, it remains unclear whether the suggested increase in dishonesty following the negative written message was elicited by an underlying mechanism – namely, a justification to re-define unethical acts as non-violating social norms, triggered by the term “forget”. This leaves us with interesting future research opportunities. For example, one could set up an experiment in which participants are given the opportunity to cheat (i.e., over-reporting their performance on an ability-based task) and thereby earn undeserved money. Prior to the task, they would be exposed to a written moral reminder framed in an affirmative or negative way, with the latter containing the term “forgetting”. After the task, participants would be asked to take a memory test regarding a list of words. Participants’ frequency of forgetting in a context where forgetting serves their interests (like in the ability-based task) would be compared to when it is preferable for participants to remember (like in the memory test). If the difference in frequency of forgetting between the two contexts would be greater for participants presented with a negative vs. affirmative reminder, then participants would most likely have used the term “forgetting” as a justification for their dishonesty.

The striking and partially unexpected results elicited by the image of a group could also be strengthened by a follow-up study. Here, one could test the possibility that the increase in dishonesty was caused by people’s impression of one specific person depicted on the image. Subtle facial features have been found to influence people’s perception of trustworthiness of their counterparts (e.g., Stirrat & Perrett, 2010). Therefore, one of the two persons that were presented in the group picture and not in the single picture might have been perceived as less trustworthy by the people exposed to the image. One way to rule out this explanation would be to use the trust game, in which different participants are asked to transfer a certain amount of money – which would be tripled – to each of the three individuals separately, who would later have the chance to

return some of it to the participants. Trust is thus measured via the amount transferred, and participants' expectations about that returned. If such amount is lower for one of the two persons depicted only in the group image as compared to the person that is also depicted in the single image, this would suggest that lower payment rates observed in the presence of a group image could be driven by people's impression of one specific face instead of the number of people depicted on an image.

Another interesting aspect to consider in future research is whether the effect of the group image replicates for more than three depicted individuals. The possible rationalization technique that employees might have used for their failure to pay (i.e., the belief that the damage to the group caused by non-accurate payments is minor) might even be amplified in the presence of a picture of, for example, eight compared to three individuals. Therefore, testing the boundaries of this effect and addressing the question whether dishonesty increases the more people that are harmed are depicted on an image seems to be a promising opportunity for future research. Furthermore, future research could examine whether people have really applied this rationalization technique by a follow-up questionnaire in which the people that were exposed to the images could be asked whether they felt that the individuals depicted on the image would be hurt by non-accurate payments.

Practical Implications

The present study suggests important practical implications that businesses should take into account to effectively convey ethical rules to their employees. By shedding light on the different effects of the wording of written moral reminders and the number of people that are harmed and depicted on an image, this study suggests a cost-effective and non-invasive way to curb dishonesty in the workplace: Framing moral reminders in affirmative ways and providing

information about one specific person harmed by dishonesty. Paradoxically, messages that highlight what not to do might lead people to do exactly that. Additionally, images depicting a group of individuals harmed by dishonesty might also lead people to increase their dishonest behaviours. Hence, businesses should be warned about the use of such ways of conveying moral standards. Instead, I advise businesses to focus on affirmatively-worded messages when conveying their moral standards to their employees and to humanize the organization by displaying one individual that is harmed by the unethical conduct.

Conclusion

Unethical behaviour in the workplace poses a major concern for corporate organizations. This research offers promising hints for how the wording of written moral reminders and the number of people suffering from the unethical behaviour that are depicted on an image can deter people from behaving dishonestly. Additionally, it highlights in which ways people's dishonesty might unintentionally be increased. Based on the results of the present study, I suggest that depicting the people harmed by dishonesty as groups (vs. single individuals) might actually increase dishonesty, and that messages highlighting what not to do might paradoxically lead people to do exactly what was asked not to be done. Thus, to promote ethical behaviour, I suggest that it is best to keep the appeal personal and affirmative.

Footnotes

¹ To ensure anonymity, I refrain from naming the delivery company that gathered the data for this thesis and the offices in which the present study was conducted.

² For technical purposes, three individuals are depicted on the group image in the present study. Since the basic number of individuals needed to form a group is two (Williams, 2010), three individuals are considered as a reasonable number. Because of limited space on the signs, it was not possible to present more than three individuals that are still well recognisable.

References

- Anand, V., Ashforth, B. E., & Joshi, M. (2004). Business as usual: The acceptance and perpetuation of corruption in organizations. *The Academy of Management Executive*, 18(2), 39-53.
- Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Journal of Personality and Social Psychology*, 74(1), 63-79.
- Bateson, M., Callow, L., Holmes, J. R., Roche, M. L. R., & Nettle, D. (2013). Do images of 'watching eyes' induce behaviour that is more pro-social or more normative? A field experiment on littering. *PloS one*, 8(12), e82055.
- Bateson, M., Robinson, R., Abayomi-Cole, T., Greenlees, J., O'Connor, A., & Nettle, D. (2015). Watching eyes on potential litter can reduce littering: evidence from two field experiments. *PeerJ*, 3, e1443.
- Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots: Why we fail to do what's right and what to do about it*. Princeton, NJ: Princeton University Press.
- Brudermann, T., Bartel, G., Fenzl, T., & Seebauer, S. (2015). Eyes on Social Norms: A Field Study on an Honor System for Newspaper Sale. *Theory and Decision*, 79(2), 285-306.
- Cai, W., Huang, X., Wu, S., & Kou, Y. (2015). Dishonest Behavior is not Affected by an Image of Watching Eyes. *Evolution and Human Behavior*, 36(2), 110-116.
- Cialdini, R. B. (2003). Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science*, 12(4), 105-109.
- Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing Social Norms for Persuasive Impact. *Social Influence*, 1(1), 3-15.

- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology*, 58(6), 1015-1026.
- Cragle, F. R. (2015, July 15). The Importance of Employee Dishonesty Coverage in the Post-Recession Era. *Hirschler Fleischer Attorneys at Law*. Retrieved from <http://www.hflaw.com/news-events/news/the-importance-of-employee-dishonest-coverage-in-the-post-recession-era>
- 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.
- Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research*, 35(3), 472-482.
- Greenberg, L., & Barling, J. (1996). Employee theft. In C. L. Cooper & D. M. Rousseau (Eds.), *Trends in Organizational Behavior* (pp. 49-64). Hoboken, NJ: John Wiley & Sons Inc.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving Persons and Groups. *Psychological Review*, 103(2), 336-355.
- Hasson, U., & Glucksberg, S. (2006). Does Understanding Negation Entail Affirmation?: An Examination of Negated Metaphors. *Journal of Pragmatics*, 38(7), 1015-1032.
- Hollinger, R. C., & Davis, J. L. (2006). Employee theft and staff dishonesty. In M. Gill (Ed.), *The Handbook of Security* (pp. 203-228). London, UK: Palgrave Macmillan.
- Jenni, K., & Loewenstein, G. (1997). Explaining the Identifiable Victim Effect. *Journal of Risk and Uncertainty*, 14(3), 235-257.

Jones, G. (2012, February 28). Good Workers Gone Bad: How to Spot Employee Theft. *CNBC*.

Retrieved from <https://www.cnbc.com/id/46556452>

Kaplan, R., & Saccuzzo, D. (2013). *Psychological assessment and theory: Creating and using psychological tests*. Melbourne, Australia: Wadsworth Cengage Learning.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The Spreading of Disorder. *Science*, *322*(5908), 1681-1685.

Keizer, K., Lindenberg, S., & Steg, L. (2011). The Reversal Effect of Prohibition Signs. *Group Processes & Intergroup Relations*, *14*(5), 681-688.

Kogut, T. & Ritov, I. (2005). The “Identified Victim” Effect: An Identified Group, or just a Single Individual? *Journal of Behavioral Decision Making*, *18*, 157-167.

Kouchaki, M., & Gino, F. (2015). Dirty deeds unwanted: The use of biased memory processes in the context of ethics. *Current Opinion in Psychology*, *6*, 82-86.

Kuratko, D. F., Hornsby, J. S., Naffziger, D. W., & Hodgetts, R. M. (2000). Crime and Small Business: An Exploratory Study of Cost and Prevention Issues in US Firms. *Journal of Small Business Management*, *38*(3), 1-13.

Leong, F. T., & Austin, J. T. (2006). *The psychology research handbook: A guide for graduate students and research assistants*. Thousand Oaks, CA: Sage.

Levy, K. J. (1980). A Monte Carlo study of analysis of covariance under violations of the assumptions of normality and equal regression slopes. *Educational and Psychological Measurement*, *40*(4), 835-840.

Maciuszek, J., & Polczyk, R. (2017). There was not, they did not: May Negation Cause the Negated Ideas to be Remembered as Existing?. *PloS one*, *12*(4), e0176452.

- Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful Negation May Depend on the Schema Used for its Encoding. *Journal of Experimental Social Psychology, 40*(4), 433-449.
- 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.
- Nettle, D., Nott, K., & Bateson, M. (2012). 'Cycle Thieves, We Are Watching You': Impact of a Simple Signage Intervention against Bicycle Theft. *PLoS one, 7*(12), e51738.
- Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical blind spots. *Psychological Science, 26*(6), 794-804.
- Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the Streets: A Field Study on Newspaper Purchasing. *Journal of the European Economic Association, 11*(3), 661-679.
- Richards, C. (1999). The Transient Effects of Limited Ethics Training. *Journal of Education for Business, 74*(6), 332-334.
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?. *Methodology, 6*(4), 147-151.
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science, 18*(5), 429-434.
- 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.
- Shannon, V. P. (2000). Norms Are What States Make of Them: The Political Psychology of Norm Violation. *International Studies Quarterly*, *44*(2), 293-316.
- Slovic, P. (2010). If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide. In S. Roeser (Eds.), *Emotions and Risky Technologies* (pp.37-59). Dordrecht, Netherlands: Springer.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349-354.
- Susskind, J., Maurer, K., Thakkar, V., Hamilton, D. L., & Sherman, J. W. (1999). Perceiving Individuals and Groups: Expectancies, Dispositional Inferences, and Causal Attributions. *Journal of Personality and Social Psychology*, *76*(2), 181-191.
- Tenbrunsel, A. E. (1995). *Justifying Unethical Behavior: The Role of Expectations of Others' Behavior and Uncertainty* (Doctoral dissertation, Northwestern University). Retrieved from <https://philpapers.org/rec/TENJUB>
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical Fading: The Role of Self-deception in Unethical Behavior. *Social Justice Research*, *17*(2), 223-236.
- Williams, K. D. (2010). Dyads can be groups (and often are). *Small Group Research*, *41*(2), 268-274.
- Yam, K. C., & Reynolds, S. J. (2016). The Effects of Victim Anonymity on Unethical Behavior. *Journal of Business Ethics*, *136*(1), 13-22.

Table 1

Means and Standard Deviations of Payment Rates During the Intervention

		Type of Picture					
		Group		N/A		Single	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Type of Message	Affirmative	.89	.22	.77	.17	.82	.11
	N/A	.71	.17	-	-	.74	.11
	Negative	.80	.30	.77	.16	.80	.08

Note. $N = 102$.

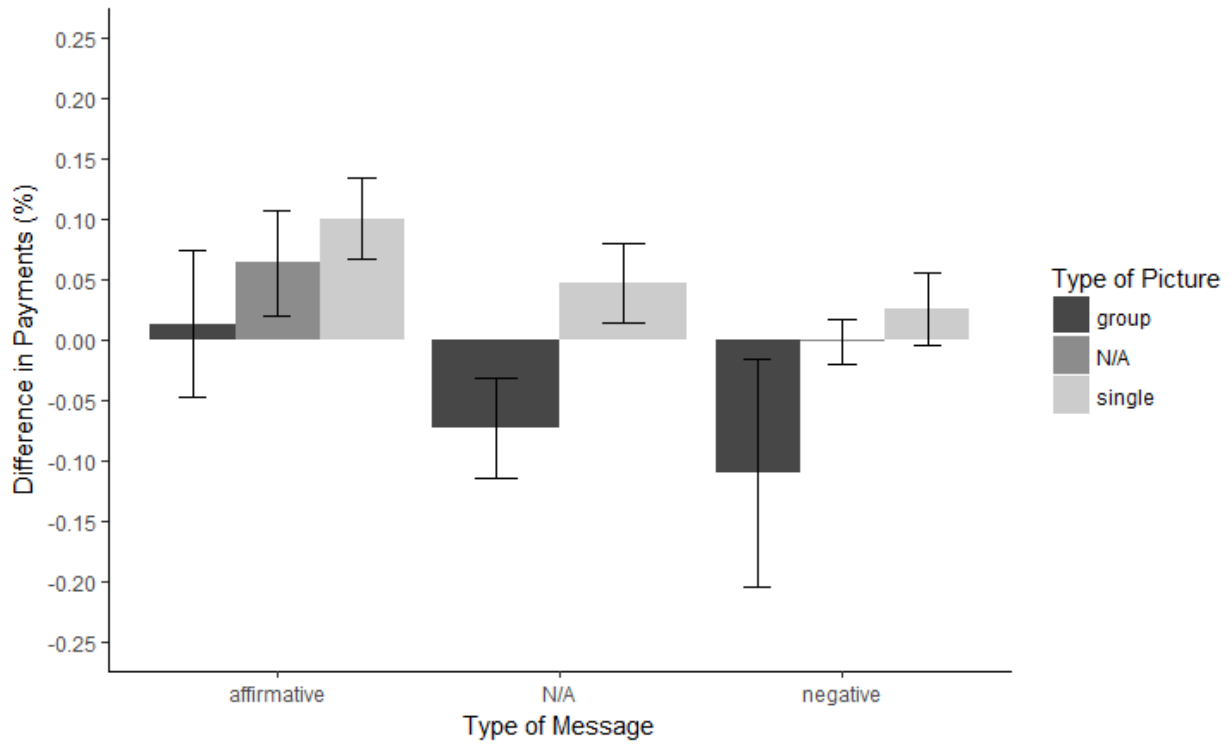


Figure 1. Differences in percentages paid during the intervention compared to the baseline level. Negative values represent decreases in payment rates, whereas positive values represent increases in payment rates. Error bars denote standard errors.

Appendix A

Signs with Written Moral Reminders

Affirmative Message

Bitte zahlen Sie hier,
wenn Sie sich einen Snack
genommen haben.

Translated: Please pay here if you took a snack

Negative Message

Bitte vergessen Sie nicht zu
zahlen, wenn Sie sich einen
Snack genommen haben.

Translated: Please do not forget to pay here if you took a snack

Signs with Images of the People that are Harmed by Dishonesty

Single Individual

Translated: Thank you from the xxx-Team,

Name, since xxx in the team

Group of Individuals

Translated: Thank you from the xxx-Team,

Name, since xxx in the team (beneath each individual)

Appendix B

Photo of a Snack Box



Figure B1. Example of a snack box presenting the image of a group of individuals combined with the affirmative written moral reminder.