

This item is the archived peer-reviewed author-version of:

Social norms of corruption in the field : social nudges on posters can help to reduce bribery

Reference:

Köbis Nils C., Troost Marleen, Brandt Cyril Owen, Soraperra Ivan.- Social norms of corruption in the field : social nudges on posters can help to reduce bribery Behavioural public policy - ISSN 2398-0648 - 6:4(2022), p. 597-624 Full text (Publisher's DOI): https://doi.org/10.1017/BPP.2019.37 To cite this reference: https://hdl.handle.net/10067/1716190151162165141

uantwerpen.be

Institutional repository IRUA

Social norms of corruption in the field –

Social nudges on posters can help to reduce bribery

*¹Nils Köbis, ¹Marleen Troost, ²Cyril Brandt & ¹Ivan Soraperra

¹Center for Research in Experimental Economics and Political Decision Making,

University of Amsterdam

²Associate Researcher, Institute of Development Policy, University of Antwerp

IN PRESS AT BEHAVIORAL PUBLIC POLICY

*Corresponding author: Nils Köbis n.c.kobis@uva.nl; n.c.kobis@gmail.com

Funding: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreements: ERC-StG-637915; ERC-AdG 295707) from Research Priority Area Behavioral Economics (University of Amsterdam, Proposal number 201806260406). The authors declare that they do not have any conflict of interests.

Acknowledgements: We thank Margarita Leib, Shaul Shalvi as well as workshop participants at the Department of Experimental Psychology at Oxford University, Department of Social Psychology at the University of Cologne, Max Planck Institute for Research on Collective Goods, and the conference participants at the 2018 Advances with Field Experiment Conference (AFE), the 2019 International Meeting series on Experimental and Behavioral Economics (IMEBESS), the 2019 conference of Subjective Probability, Utility, and Decision Making (SPUDM) for useful comments. We also thank Marijke Bezuidenhout, Phumzani Ntuli and Phumlani Ngwenya for their help with data collection.

Abstract

Corruption in the form of bribery continues to be a major societal challenge around the world. The current lab-in-the-field study tested whether descriptive norms messages on posters can help to reduce bribery. Before, during and after placing posters throughout a medium sized South African town, incentivized measures of social norms and bribery were assessed in a mobile lab. A total of 311 participants stemming from the general population took part. In line with the preregistered predictions, the results reveal (a) that people perceive bribery to be less common and (b) engage in bribery in a corruption game less frequently when the posters were displayed. The discussion outlines how social norms nudging campaigns can be leveraged to spur collective action against corruption.

Keywords: corruption; bribery; social norms; descriptive norms; social nudging

Social norms of corruption in the field – Social nudges on posters can help to reduce bribery

Corruption continues to produce profoundly negative consequences for societies around the world (Rothstein & Varraich 2017). Massive anti-corruption efforts have largely failed to reduce daily forms of corruption such as bribery (Fisman & Golden 2017). The failure stems in part from traditional anti-corruption approaches that underestimated the social embeddedness of systemic corruption (Rothstein 2011) and a general neglect of behavioral features of corruption on the individual level (Heywood 2018). Instead, social norms accounts combining social and individual factors have gained popularity as a theoretical framework to study corruption (Kubbe & Engelbert 2018).

Prominent social norms frameworks distinguish between an injunctive and a descriptive element of social norms (Goffman 1969; Paprzycka 1999; Bicchieri 2016; Cialdini et al. 1990). The injunctive element conveys what is considered as acceptable, while the descriptive element refers to what is considered as common (Tankard & Paluck 2016). Recent theories on corruption highlight the descriptive element of social norms as a crucial predictor for corrupt behavior (Stephenson 2018; Rothstein 2011; Marquette & Peiffer 2015). For example, current theories on corruption suggest that people in high corruption contexts often engage in petty forms of corruption because they believe that others around them do so too – even though they themselves might consider it wrong and unacceptable (Rothstein 2000).

Evidence from controlled lab experiments suggest that social nudges targeted at changing perceived social norms can reduce bribery (Köbis et al. 2015). However, field evidence for such social norms campaigns to reduce bribery is lacking – as recent systematic literature reviews on social norms of corruption reveal no studies that test social norms as anti-corruption tool in the field (Köbis et al. 2019). Could similar social nudges also transform people's beliefs and reduce

bribery in the field, where people are exposed to a multiverse of information in their environment? In pursuit of answers, we conducted a <u>pre-registered</u>, financially incentivized labin-the-field experiment in a country struck by high levels of corruption: South Africa. Our results provide first promising insights illustrating changes in perceived descriptive norms and bribery in the corruption game.

Theories on the persistence of corruption

Corruption is widely defined as the "abuse of public office for private gain" (Nye 1967; Rose-Ackerman 2006). The interdisciplinary literature unanimously describes stark differences in corruption levels around the globe (Heywood 2014). Influential theoretical accounts have formalized these global differences in frequency-dependent equilibria models (Andvig & Moene 1990; Bardhan 1997). These models outline that societies either rest in a state of low or high corruption. To illustrate this view, consider the classical example of bribing a police officer to avoid a fine for a minor traffic violation. In a country where bribery rarely occurs (=low corruption equilibrium), offering a bribe likely leads to bigger trouble than paying the fine. The reason being the danger of the police officer reporting and punishing bribe attempts. However, in a context where bribery is commonplace (=high corruption equilibrium), paying bribes represents the "best choice" to get ahead. Here, corruption is theorized to be a social trap meaning that once corruption has become systemic, it tends to reinforce itself (Stephenson 2018). One major reason stems from the fact that legal enforcement institutions themselves fall prey to the high levels of corruption (Persson et al. 2012). In the absence of reliable punishment, corruption turns into a collective action problem (Rothstein 2000).

In the corrupt practice of bribery, for each individual, the decision whether to bribe or not to bribe reflects a social dilemma: a conflict between short-term self-interest (=reaping the benefits of bribery) and long term collective interest (=corruption free society) (Köbis et al. 2018; Köbis et al. 2016). As in other social dilemma type situations (Kerr & Kaufman-Gilliland 1997; Bicchieri & Dimant 2019), the expected benefit from bribery hinges on the belief about the behavior of others. In short, the collective action problem of bribery consists of people bribing as they think it is common, even though they might not approve it.

Towards a social norms perspective of corruption

Social norms accounts can capture these dynamics and have therefore received increased attention among scholars seeking to understand and change socially entrenched practices (Efferson & Vogt 2018), also among corruption researchers (Kubbe & Engelbert 2018) and for anti-corruption policy (Jackson & Köbis 2018; Scharbatke-Church & Chigas 2016). That is, using social norms as an analytical lens allows to study the occurrence of a particular corrupt practice by combining social and individual factors (Cislaghi & Heise 2018; Bicchieri 2016). Particularly, frameworks distinguishing between injunctive (=acceptability) and descriptive (=frequency) norms are well suited to analyze the collective action problem of corruption, as shown in recent lab studies (Abbink et al. 2018; Köbis et al. 2015; Zhao et al. 2017; Schram et al. 2019).

The perceptions of social norms are however subject to distortions (Perkins & Wechsler 1996). In fact, people often overestimate actual levels of corruption, in part due to its secretive nature which often prevents an actual observation of corrupt practices (Pinker et al. 2008; Mény 1996). In many societies, narratives about corruption exists signifying that "everybody does it" which further perpetuates the exaggerated perceived descriptive norms about corruption – and eventually sustains the outlined social trap of corruption. As many traditional anti-corruption

policies have largely failed to escape this corruption trap (Mungiu-Pippidi 2017), new hope has been placed in behavioural approaches to fight corruption, most importantly (social) nudging.

Behavioural approaches to reduce corruption: Social norm nudging

Applying behavioural economics and (social) psychology to public policy has increased over the last decades (Camerer et al. 2003; Thaler & Sunstein 2008; Oliver 2015). Such behavioural approaches are also gaining increasing popularity to fight corruption (see for example, OECD 2018). Although behavioural approaches can contribute in manifold ways to public policy (Oliver 2013; Loewenstein & Chatter 2017), here we focus on the most popular behavioural policy tool namely nudging. According to Thaler and Sunstein (2008) nudging refers to "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (p.6; Thaler & Sunstein 2008).

More specifically, we zero in on social norm nudging, defined as "a nudge whose mechanism of action relies on social norms, i.e., on providing social information or eliciting social expectations with the intent of inducing desirable behaviour" (p. 2; Bicchieri & Dimant 2019). Norm nudging campaigns are based on the outlined social norms theories in assuming that people's decisions are often subject to social influences. When crafted carefully, such norm nudging can prove successful to change behaviour, as evidenced by a large corpus of studies across a wide array of practices (see for a meta-analysis, Manning 2009).

When it comes to corrupt conduct, however, the insights so far merely rely on lab studies. For example, findings from Barr and Serra (2010) suggest that perceived descriptive norms about corruption are indeed malleable. Other lab studies suggest that providing social information about low levels of bribery among peers reduces people's own propensity to bribe – hence provided first causal evidence running from descriptive norms to corrupt behaviour (Köbis et al. 2015). As an extension of these findings, we tested the effectiveness of norms nudges as a vehicle to reduce corruption in the field.

We distributed posters with descriptive norms messages about bribery throughout Manguzi, a medium sized town in the region of KwaZulu-Natal (South Africa). Perceived levels of corruption have remained high throughout South Africa, also in the past-apartheid governments, fortifying socio-economic inequalities within the country (Koelble 2018; Peiffer et al. 2019). However, survey data from South Africa including the Southern region of KwaZulu-Natal shows a positive trend – with self-reported bribery frequencies decreasing over recent years (Transparency International 2013,2017, Ethics Institute 2016, 2017). Making use of this reduction, we displayed information about the decreasing levels of bribery on the posters.

We are by no means the first to attempt to reduce corruption through information campaigns (Johnsøn et al. 2012; Peiffer & Walton 2019). However, we delineate from previous approaches in two main ways: First, instead of trying to increase awareness about the negative impact of corruption, we use a social norms approach targeting the beliefs about the behavior of others. More precisely, to avoid backfiring effects of social nudges (Bicchieri & Dimant 2019), the message displayed trends in descriptive norms (see also Mortensen et al., 2017). Hence, instead of static information (e.g. "12 percent of South Africans bribe") the poster conveyed dynamic trends ("Less and less South Africans bribe"). As a second main difference to previous information campaigns, we evaluated the intervention with incentivized behavioral data within a mobile lab erected in the field. Therein, we assessed social norms and corrupt behaviour with a novel bribery game. We tested the following pre-registered hypotheses: H1: Perceived descriptive social norms about bribery are lower during the Poster treatment compared to the Baseline treatment.

H2: Perceived injunctive social norms about bribery do not statistically differ between both treatments.

H3: The levels of bribery in the bribery game are lower during the Poster treatment compared to the Baseline treatment.

Method

Incentivized norms assessment

We used the incentivized norms elicitation method introduced by Krupka and Weber (2013). This method consists of scenarios in which 'Person A' has various choices and participants judge the frequency and acceptability of a given decision. Participants receive financial rewards for accuracy, i.e. when their evaluation matches that of another randomly selected participant. Hence, incentives exist to reveal their actual belief about the collectively shared view about frequency and appropriateness as opposed to their own personal judgement. The method thus consists of a coordination game to assess social norms. While incentivized norms assessment methods require more substantive explanation and instructions, they have the advantage of overcoming some of the challenges that non-incentivized measures of social norms face, such as social desirability concerns (Bicchieri et al. 2014).

The current study contained three such scenarios, one of them describing a bribery transaction. For each scenario, we used three social norms items to measure both descriptive and injunctive norms: First, to assess descriptive norms participants indicated their beliefs about the frequency of a given course of action. Second, to measure injunctive norms participants indicated their beliefs about the social appropriateness of this behaviour. For both questions we matched a participant's response to that of another randomly selected participant. Providing the same answer earned both participants R5 ($\approx \in 0.3$). As a third question, we asked participant for their personal appropriateness rating of the behaviour. For this question, we did not match participants' answers and provided no incentives as it measured participants' private perception. Instructions on the screen extensively explained this procedure and two test questions ensured participants' understanding (see all material on OSF).

Social dilemma bribery game

We used a corruption game that models bribery as a social dilemma to reflect the psychological decision structure existent in highly corrupt contexts. The game consists of a 10-player game in which participants take the role of 'citizens' and 'public officials'. Reflecting commonly occurring transactions between citizens and public officials, citizens seek to obtain a certificate, e.g., a driving license, that is valuable to them (R35). A citizen can either apply for the certificate by paying the regular application fee of R15 or can pay a bribe of R10 to avoid the application fee. The public official equally has two options: either process the application and earn a wage of R20 or earn additional money by accepting the bribe of R10. Importantly, each successful bribery transaction incurs a loss of R2 on all 10 players, reflecting the social cost of bribery (see Figure 1). The payoffs are outlined in more detail below.

Figure 1 about here

Payoffs for the citizen. The earnings of the citizen depend on three factors. First, the citizen's own decision to pay the application fee or bribe. Second, when opting to bribe, the payoff depends on whether the public official accepts the bribe. Third, the citizen's payoff also depends on all other players in the matching group, i.e. whether the other citizen/public official pairs successfully coordinate on bribing which results in a social cost of R2 on all the 10 players. If the public official accepts, the citizen receives the certificate without having to pay the application fee, at the same time such successful bribery incurs a cost on all players in the matching group. Therefore the citizen's final payoff is:

R35 - R10 - R2 - S

Where S denotes the social cost of corruption generated by the other pairs belonging to the matching group, i.e., S = R2 * Number of bribes accepted in the other pairs of the matching group. If instead the public official rejects the bribe, the citizen has to pay the application fee of R15 and an additional cost of R5 resulting in a payoff of:

Finally, if the citizen decides not to offer the bribe, the citizen has to pay the application fee of R15 resulting in a payoff of:

R35 - R15 - S

Payoffs for the public official. The public official's payoffs depend on the same three factors influencing the citizen's payoff. First, on whether the matched citizen offered a bribe or not. Second, on the public official's own decision to accept or reject the bribe. Third, on the successful bribery transactions by other citizen/public official pairs. If the citizen does not offer a bribe the public official earns the wage of R20 and his final payoff is:

R20 - S

If the public official instead is offered a bribe and accepts the offer, the payoff is:

R20 + R10 - R2 - S

Finally, if the official is offered a bribe and rejects the offer, the payoff is:

R20 - S

Whatever the behaviour of the other pairs in the matching group, the public official has an incentive to accept the bribe when offered one. In this way the public official increases the payoff by R8. Anticipating this, the citizen has an incentive to offer the bribe. In this case, the citizen can increase the payoff by R3. Engaging in corruption thus represents a dominant strategy while leading to socially worse outcomes, a social dilemma (Köbis et al. 2016). Indeed, if all 5 pairs refrain from corruption the payoff of both citizens and public officials is R20. If instead all the pairs engage in corruption, the public officials obtain R20 but the citizens obtain only R15. These relatively small amounts reflect the notion that the majority of bribery transactions in KwaZulu-Natal are petty in nature, as recent survey evidence suggests that the majority of the amounts paid for bribes do not exceed 100R (~ ϵ 6,40; see Dobie, 2017). Moreover, we deliberately chose not to include punishment in the game in order to reflect that the likelihood of being caught and prosecuted for petty forms of bribery is very low – if not absent in South Africa (Peiffer et al. 2019).

We used the strategy method (Brandts & Charness 2011) in which each player makes a decision as a citizen and as a public official. In matching groups of 10, one decision per participant is randomly chosen and matched to that of another participant. We use a lightly framed version of the game, which means that we used the labels 'citizen' and 'public official' but referred to the bribe as a 'side payment' to reduce social desirability concerns. To ensure that participants understood the logic of the game we displayed figures illustrating the structure of the game (see Figure 1) and assessed participants understanding using three test questions prior to the game.

Additional Measures

Besides standard demographic information (age, gender, education), we included three control questions. First, we assessed at which day of the month participants receive their salary. This payment differs across professions and indicated the sector that the participant worked in and whether the participant was employed at all. Second, we examined whether participants receil seeing the poster. We asked participants if they noticed a poster that reminded them of the topic of the study (Y/N), and if yes, whether they can recall any of the words mentioned on the

poster. We also included a one-item measure of perceived corruptness of the side payment ("How corrupt do you think the transaction was?") with answers given on a 7-point scale ranging from "not corrupt at all" to "very corrupt".

Payment method

In addition to a show-up fee of R25 ($\approx \in 1.50$), participants received between R25 and R45 ($\approx \in 1.50$ to $\in 2.73$) for completing the study which on average took 27.6 minutes. These earnings corresponded to 2-5 hourly wages based on current minimum wage levels for the Extended Public Works Programme (R11/hour). Participants received their payment within 48 hours after completion of the study via Instant Money Transaction. It sends a voucher together with a PIN to a cell phone number via SMS. Participants could redeem this voucher without any identification at any ATM or Money Market counter, hence ensuring privacy. We deleted all cell phone numbers upon successful payment. The payment method did not allow to pay precise amounts (e.g. R23.50). To determine whether the final pay would be rounded up or downwards, we let participants roll and report a die (for more details on this, see Table S1, in SOM).

Treatments in the experiment

The experiment consisted of two treatments: First, during the Baseline treatment Research Assistant 1 (RA1; blind to the hypotheses of the study) conducted the lab-in-the-field experiment as outlined above. Data for the Baseline treatment was collected in two waves: from 16/07/2018 to 28/07/2018 (week 1) and from 05/09/2018 to 06/09/2018 (week 7). We initially scheduled this baseline treatment for the first week of the study (prior to the poster treatment) and in the fourth week (after the poster treatment). However, due to unforeseen difficulties in the data collection we had to postpone this second assessment to the seventh week which we separately pre-registered prior to data collection (see <u>Amendment to pre-registration in SOM</u>). Second, during the Poster treatment Research Assistant 2 (RA2, blind to the purpose of the study) put up 20 posters reading the message "Less and less people from KwaZulu-Natal pay bribes" (see Poster in Figure 2). The factual validity of this statement stems from large national surveys (see Transparency International 2013, 2017; Ethics Institute 2016, 2017), which show a decrease in self-reported bribery. RA2 placed the posters in several locations throughout town, mostly on signposts next to the road as well as in other places where pamphlets are often displayed. The supplementary material shows the exact locations on a map and contains pictures of the poster sites (see <u>OSF</u>). The posters were also put up at the entrance of the building in which the mobile lab was installed. It is important to note that the poster were not on display within the lab, including the waiting area. Data for the Poster treatment was collected from 30/07/2018 to 04/08/2018 (week 2). We determined the sample size before any data analyses (see <u>Pre-Registration</u> for a priori power calculations).

Figure 2 about here

Results

Overall, 311 participants took part in the experiment (Baseline: N = 187; Treatment: N = 124). As a sign of successful randomization, we find no difference between treatment and baseline groups across main demographic variables (see Table S2 in the SOM). Moreover, the two waves of the Baseline treatment (pre and post poster) are largely equivalent as well – merely differing with regards to participants in the second wave being slightly younger, having obtained higher levels of education and indicating slightly more polarized views about the social injunctive norm about bribery (see full overview in Table S3 in the SOM).

An unforeseen anomaly occurred in the data collected on the 30th of July. After completing the study, one participant revealed to RA1 that most of the subjects participating in that day coordinated their responses to the descriptive norms item prior to the experiment. They were peers working in the same office and gathered information about the most profitable answers to the descriptive norm questions in the past week. That means, participants coordinated on the most common answer during the data collection of the Baseline treatment. In fact, 21 out of 24 subjects participating in the experiment on the 30th of July gave exactly that same answer in the descriptive norm question, earning them the maximum payoffs for that task (see Figure S1 in SOM). Given this irregularity, we control for data stemming from that day using a dummy variable in the analyses, below referred to as collusion dummy.

Descriptive and Injunctive norms

Overall, we observe that the majority of respondents believe that most people pay bribes as 60.77 percent indicate that more than 70 percent of people pay bribes. This pattern is robust to controlling for the collusion dummy 57.49 percent of respondents believing that bribery is higher than 70 percent. This pattern supports the basic premise of a coordination in that the majority of people think that the majority of people bribe.

To test H1 predicting that these perceptions would decrease during the poster treatment, we conducted ordered probit regression analyses predicting the incentivized descriptive norm variable. With our data, i.e., categorical dependent variable that can be ordered, this regression technique is better suited than linear regression analysis because it does not assume the same distance between categories. The ordered probit regression permits to estimate whether there is a directional shift in the probability to observe an answer from lower to higher categories, while controlling for other explanatory variables. Controlling for the data collection anomaly, we observe a significant shift towards lower perceived descriptive norms, both when only including the treatment dummy (Model 1, Table1) and when additionally controlling for demographic characteristics (Model 2, Table 1). Hence, we find evidence that the poster reduced perceived descriptive norms about bribery, confirming the first hypothesis (see also Figure 3).

Table 1 & Figure 3 about here

Next, we report descriptive statistics for both injunctive norms items (social and individual). First, a majority of participants (67.52%) believe that others consider bribing as (very) socially inappropriate (68.99% when controlling for collusion). Second, the majority of participants (66.24%) also considered bribery themselves as (very) socially inappropriate (67.94% when controlling for collusion). Both results suggest that a majority of people collectively and individually disapprove of bribery, even in corrupt societies.

To test H2 predicting no shifts in both injunctive norms items across both treatments, we again conducted ordered probit regression analyses. Table 1 depicts the results for social and individual injunctive norms. In support of H2, the findings reveal no statistically significant directional shift in the frequency of answers neither without (Models 3 & 5), nor with controlling for demographic characteristics (Models 4 & 6). There is, instead, a reduction in the frequency of the answer "Somewhat socially inappropriate" in favor of all the other answers. the parameter of the treatment dummy is not significantly different from zero both.

As an additional (non-pre-registered) test for potential shifts towards higher social inappropriateness of bribery during the poster treatment we conducted Bayesian analysis. For this analysis, we dichotomized the social and individual injunctive norms items into a dummy of 0 (= (very) inappropriate) and 1 (= (very) appropriate). The results reveal Bayes factors of $BF_{01} > 18.84$ for the social injunctive norms and $BF_{01} > 11.99$ for the individual injunctive norms. ¹ Together, these results constitute strong evidence (Jeffreys, 1961) that the null hypothesis of no shift is more likely than an alternative hypothesis that predicts a shift towards higher social inappropriateness during the poster treatment. Overall, the results confirm H2 and provide convergent validity for the poster treatment specifically targeting descriptive norms.

Bribery behaviour

Next we test H3, predicting that bribery levels in the corruption game drop during the Poster treatment. As Figure 4 illustrates, the percentage of subjects offering (panel a) and accepting bribes (panel b) decreases during the Poster treatment compared the Baseline treatment. To test the significance of these differences, we conducted binary regression analyses with robust standard errors for the day of data collection (see Table 2). For the citizens' bribe offers the reduction in bribery during the poster treatment is not significant, neither when tested alone (Model 1), nor when controlling for demographic variables (Model 2).

For the public officials' bribe acceptance, the reduction of bribery during the poster treatment is p = .062 when analyzed alone (Model 3) and p < .001 when controlling for demographic variables (Model 4). Hence, although the reduction of bribe offers in the role of citizens is not statistically significant, we do find some support that is statistically significant for a reduction in bribe acceptance in the role of public officials.

¹ We conducted the Bayesian analysis in JASP (<u>https://jasp-stats.org/</u>) for contingency tables under the assumption of independent multinomial sampling and used the number of participants per treatment as fixed margin (see for more details, see Jamil et al. 2017). The results for the social injunctive norms indicate a Bayes Factor of BF01 = 28.89 (BF01 = 18.84 for the subset without collusion) and for the personal injunctive norms a Bayes Factor of BF01 = 19.41 (BF01 = 11.99 for the subset without collusion).

Table 2 about here

Finally, Figure 5 illustrates the association between the decisions made by the participants in the role of citizens and public officials. The choice pattern reveals a high degree of consistency in the choices across both roles, with the majority of the participants either opting for or against bribery in both roles. Overall, we find a higher propensity towards bribery when acting as public official compared to when acting as citizen (see higher percentages in 'accept' column vs. in the 'offer' row in Figure 5). Moreover, McNemar's test reveals a higher propensity towards bribery when deciding as a public official vs. citizen in the Baseline treatment (p < .001). In the Poster treatment the higher propensity for bribery as a public official (vs. citizen) is p = .093 – again indicating a reduction of public officials willing to accept bribes in the Poster treatment. Overall, we find a decrease in unconditional bribery (accept/offer) of 5.8 percentage points and an increase in unconditional rejection of bribery (don't offer/don't accept) of 9.5 percentage points.

Recognition of the poster

Analyzing the question whether participants have recently seen a poster related to the topics of the survey, we observe that 39.6% (72/182) declared to have seen a poster in the Baseline and 70% (70/100) declared to have seen a poster in the Poster treatment. Of the participant that declared to have seen a poster when our poster was present, 92.9% (65/70) were able to recall the poster content. Results for analysis without the subjects that did not recall the content of the poster when our poster are reported in Tables S6 & Table S7 in the SOM.

Discussion

The results provide first behavioural insights into the effectiveness of descriptive norms information as a vehicle for anti-corruption in the field. First, the norms nudge on posters reduced people's own perceived descriptive norms about bribery, while leaving their perceived injunctive norms unaffected. Second, the poster treatment also affected participants' willingness to accept bribes in an incentivized bribery game – providing first indications that norms trend messages outside of the lab can reduce both perceptions and behaviour of corruption inside the mobile lab. We discuss the findings in light of their relevance for theory and policy on social norms of corruption.

From a theoretical perspective, the results support a "corruption trap" view (Stephenson 2018). For one, our results show that people perceived bribery as commonplace. This insight confirms previous views that in Manguzi (South Africa), side payments such as the one modelled in our bribery game are often considered the standard way to operate (Gitau, 2011). At the same time, we find first evidence using incentivized norms measures that people consider bribery as inappropriate – both personally and socially. It thus appears that people often bribe not because they consider it the right thing to do but rather because they have little other options due to the perceived high frequency of bribery around them (Fisman & Golden 2017; Bardhan 1997; Rothstein 2000; Köbis et al. 2018).

The findings also confirm previous theorizing postulating that when descriptive and injunctive norms are incongruent, descriptive norms often exert a stronger pull (Bicchieri & Dimant 2019). Here, we replicate previous findings from the lab (Köbis et al., 2015), showing that indeed the descriptive element seems to trump the injunctive element of social norms when it comes to the decision to engage in bribery. Importantly, we note that the findings neither

suggest that injunctive norms do not matter at all, nor that descriptive norms necessarily overshadow injunctive norms for other corrupt practices or in other social contexts. Instead, we argue that distinguishing between descriptive and injunctive norms provides a useful framework to investigate the respective social and behavioural elements of corruption in a particular context (Hoffmann & Patel 2017; Heywood 2017).

The findings also contribute to the growing literature on norm nudges – providing insights into how they can be used for anti-corruption. For one, by recognizing that not one monolithic corruption norm exist in a given society, anti-corruption efforts can be tailored to the social normative pressures of a particular corrupt practice. Prominent theories on social norms emphasize the importance of identifying the relevant reference network, i.e. the "range of people whom we care about when making particular decisions" (p. 14; Bicchieri 2016). Here, cognizant of the importance to tailor norms messages to the relevant reference network, we opted to refer to people from KwaZulu-Natal, rather than South Africa. Such approaches could be extended by messages that target even more specific reference networks from which social normative pressures around corruption emanate. For example, whether public officials extract side payments also depends on the normative pressures on the office floor, both stemming horizontally from peers and vertically from superiors (Jackson & Köbis 2018). Moreover, we focus both in the poster and in the design of the game on bribery because it entails an interactive form of corruption in which beliefs about the counterpart are of particular importance. Future efforts could examine whether similar information campaigns also have an effect on less interactive forms of corruption, such as embezzlement

We do note, however, that merely relying on norm nudges to fight the deeply rooted behaviour of corruption does not suffice. Akin to other domains such as obesity (Oliver & Ubel 2014), also for anti-corruption, nudges alone unlikely solve the problem as behavioural approaches in general should not be seen as a substitute but rather a supplement to traditional (anti-corruption) policies (Loewenstein & Chatter 2017). Recognizing the importance of social norms to explain corrupt practices is essential for more traditional anti-corruption approaches to succeed. For example, the relative ineffectiveness of many public salary increase programs (see systematic literature review reported in Soraperra et al. 2019) in part stems from the persistence of social norms around corruption. A case in point: a salary increase program within the tax administration in Uganda backfired, because increased wages also increased social normative expectations for public officials to provide for their family – leading public officials to extract *more* bribes (Baez-Camargo 2017; Fjeldstad 2005).

Policies for lasting social norms change of corruption

Our intervention targeting social norms revealed that perceived social norms and bribery inclinations regressed back to pre-treatment levels – hence the effect was temporary, which is not uncommon for nudging interventions (Frey & Rogers 2014). To increase the chances for enduring reductions of corruption, social norms messaging campaigns should be accompanied by complementary polies. Based on previous work on social norms change, particularly promising approaches consist of: (a) community involvement, (b) information campaigns via other media channels, and (c) social network analysis. Let us briefly address some of the implications of our study for each of these approaches.

First, community meetings featuring discussions about normative practices can contribute to cascades of social norms change (Bicchieri & Mercier 2014) – an approach that has already been successfully tested in other domains such as voting behaviour (Banerjee et al. 2011). The individual and social disapproval of corruption observed in the current study could facilitate the collective action against bribery as it provides a breeding ground for the creation of new norms, especially if these norms are observably implemented by peers (Dimant 2015).

Second, complementary approaches could use the immense anti-corruption potential of (free) media (Starke et al. 2016). In particular, radio programs, TV shows like soap operas, or even children's anti-corruption books could help to provide new narratives around social norms of corruption (see for examples, Pailey 2013; Pailey 2019). These measures have been evaluated in various domains showing promising potential to even change culturally embedded and hidden collective practices such as female genital cutting (Vogt et al. 2016; Paluck 2009) - yet, empirical evidence for their effectiveness to curb corruption is lacking.

Third, to unleash the full potential of localized social norms campaigns such efforts could be combined with social network approaches (Tankard & Paluck 2016). Mapping and tracking information dispersion within a given social network has been proposed in recent anti-corruption policy work (Jackson & Köbis 2018; Hoffmann & Patel 2017) as it has shown promising potential in other domains such as high-school bullying (Paluck et al. 2016). In sum, social nudges alone do not suffice to curb corruption, yet they can serve as an important complementary step to accompany "heavy lift" policies, such as structural reforms (Sunstein 2015; Loewenstein & Chatter 2017).

Evidence-based anti-corruption

For the design of research evaluating these and other behavioural interventions to fight corruption, several points deserve mention. First, as a method for evaluating anti-corruption policies, our study presents the use of a lab-in-the-field design with a behavioural measure of corruption. Obtaining "hard" evidence for the reduction of corruption represents a thorny challenge due to its secretive nature and the high social desirability concerns (Olken 2009). Reviewing field studies on the effectiveness of information campaigns for anti-corruption, Winters, Testa and Fredrickson in fact state that the majority of studies do not use corruption as the dependent variable but "rather some observable behaviour that is assumed to affect corruption" (p. 234, Winters et al. 2012). Such indirect measures can be problematic when trying to understand the actual corrupt practices. We propose that behavioural measures such as bribery games can serve as tools to pilot test and evaluate interventions, and thus contribute to evidencebased anti-corruption (Mungiu-Pippidi 2017).

Second, we introduce a new game that models the bribery transactions as a social dilemma. Based on recent theoretical developments in behavioural science, we argue that such a game reflects the actual psychological decision making process in highly corrupt context more accurately than many other corruption games (Köbis et al., 2016). Although bribing in the game surely differs in multiple ways from the "real world", we argue that the data obtained here bears merit: for one, because participants did perceive the side-payment as corrupt, underlining the internal validity of the measure (see Table S11, SOM). Furthermore, previous research using bribery games, shows a large overlap between corrupt behaviour in the lab and in the field, providing external validity (Armantier & Boly 2012).

Third, we pre-registered the experiment in which we outlined the data collection and analysis strategy. However, several unforeseen occurrences led us to deviate from this preregistration plan (see complete list in SOM). Most importantly, we had to include an additional control variable due to an unforeseen coordination scheme that took place on the first day of the poster treatment. Although we instructed the research assistant to take measures to prevent such schemes from occurring and no further evidence of similar incidences exist, we cannot exclude the possibility that also during other days participants coordinated on responses prior to the study. While we argue that controlling for this day of data collection is justified in order to get the most empirically valid answer to our key questions, the occurrence of this scheme represents a remarkable finding in and of itself.

Finally, given that our study was conducted over a relatively short time frame in a single location, a logical candidate to overcome some of the limitations is to run a full-fledged Randomized Control Trial (RCT). Although RCTs can indeed potentially provide valuable new insights into the workings of social norms of corruption in the field, it is important for follow-up research to consider several famously voiced drawbacks of that approach, when seeking to evaluate anti-corruption policies (Bédécarrats et al. 2019; Deaton 2010). The first main point of criticism posits that RCTs often narrowly focus on short-term impact (Bédécarrats et al. 2019). To overcome this limitation and to test the longevity of interventions, we see particular potential in adopting a long-term perspective. The use of longitudinal methods could allow to evaluate existing approaches to change social norms around corruption, such as whether new narratives in public media can indeed change norms around corruption (Pailey 2019). As a second point of criticism, many RCTs neglect mechanisms and processes that drive the effect of an intervention (Bédécarrats et al. 2019). In light of the importance of dual-process models to explain the mechanisms of nudging (Oliver 2013) and (un)ethical behaviour (Köbis et al., 2019), gaining new insights into the cognitive processes of norm nudging appears especially fruitful. That is, future work could address whether the mechanism behind belief updating observed here primarily stems from an intuitive or a deliberative process. As a final critical point, RCTs "are subject to a host of political influences as much in their design and execution as in the dissemination of their results."(p. 10; Bédécarrats et al. 2019), making upscaling a challenging matter. Also, research on nudging campaigns shows that the messenger of the social norms

information can influence the effectiveness of the intervention. Upscaling a norm nudging campaign via government institutions could thus potentially face the challenge that the messenger is not trusted (see for example, Stibe & Cugelman 2016) – put differently people might reject an official norms message campaign because it is an *official* norms message campaign (Arad & Rubinstein 2015). When taking these recently voiced considerations into account, future policy-oriented research could thereby provide further relevant insights for anti-corruption approaches building on behavioural insights.

Conclusion

Behavioural interventions using social nudges have traditionally steered away from tackling "some of the world's biggest and most complex social dilemmas" (p. 211; van der Linden 2018). To fill this gap, we conducted a lab-in-the-field experiment that took on a daunting task: changing people's perceptions about corruption and subsequently lowering their willingness to engage in bribery with the mere distribution of posters. In light of the deeply socially entrenched character of systemic corruption and the fact that people form beliefs about corrupt practices through a multitude of channels (media, own experience, social interactions), it is a striking observation that a descriptive norms poster successfully changed the incentivized social perceptions and people's willingness to bribe in the subsequent game played for real money. By providing field evidence for effectiveness of informational campaigns to change beliefs and behaviour, the results could potentially mark one first step towards an escape path from the corruption trap and illustrate the importance of behavioural approaches to anti-corruption.

References

- Abbink, K. et al., 2018. The effect of social norms on bribe offers. *The Journal of Law, Economics, and Organization*, 34(3), pp.457–474.
- Andvig, J.C. & Moene, K.O., 1990. How corruption may corrupt. *Journal of Economic Behavior* & Organization, 13, pp.63–76.
- Arad, A. & Rubinstein, A., 2015. The People's Perspective on Libertarian-Paternalistic Policies, Available at: http://arielrubinstein.tau.ac.il/papers/LP.pdf.
- Armantier, O. & Boly, A., 2012. On the external validity of laboratory experiments on corruption. *Research in Experimental Economics*, 15, pp.117–144.
- Baez-Camargo, C., 2017. Corruption, Social Norms and Behaviours: A comparative assessment of Rwanda, Tanzania and Uganda, Basel, Switzerland: Basel Institute of Governance.
- Banerjee, A. V et al., 2011. Do informed voters make better choices? Experimental evidence from urban India. *Journal of Development Studies*, (nov), p.46.
- Bardhan, P., 1997. Corruption and development: A review of issues. *Journal of Economic Literature*, 35, pp.1320–1346.
- Barr, A. & Serra, D., 2010. Corruption and culture: An experimental analysis. *Journal of Public Economics*, 94(11–12), pp.862–869.
- Bédécarrats, F., Guérin, I. & Roubaud, F., 2019. All that Glitters is not Gold. The Political Economy of Randomized Evaluations in Development. *Development and Change*, 50(3), pp.735–762.
- Bicchieri, C., 2016. Norms in the Wild: How to Diagnose, Measure, and Change Social Norms 1 edition., New York, NY: Oxford University Press.
- Bicchieri, C. & Dimant, E., 2019. Nudging with Care: The Risks and Benefits of Social

Information. Public Choice, Forthcoming..

- Bicchieri, C., Lindemans, J.W. & Jiang, T., 2014. A structured approach to a diagnostic of collective practices. *Frontiers in Psychology*, 5, p.1418.
- Bicchieri, C. & Mercier, H., 2014. Norms and Beliefs: How Change Occurs. *The Jerusalem Philosophical Quarterly*, 63(January 2014), pp.60–82.
- Brandts, J. & Charness, G., 2011. The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, 14(3), pp.375–398.
- Camerer, C. et al., 2003. Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism." *University of Pennsylvania Law Review*, 151(3), p.1211.
- Cialdini, R.B., Reno, R.R. & Kallgren, C., 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), pp.1015–1026.
- Cislaghi, B. & Heise, L., 2018. Using social norms theory for health promotion in low-income countries. *Health Promotion International*, pp.1–8.
- Deaton, A., 2010. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2), pp.424–455.
- Dimant, E., 2015. On Peer Effects: Behavioral Contagion of (Un)Ethical Behavior and the Role of Social Identity. *Journal of Economic Psychology*, 73, pp.66–88.

Dobie, K., 2017. South African Citizens' Bribery Survey, Pretoria, South Africa.

Ethics Institute, 2016. South African Citizens' Bribery Survey, 2016 (Report no. 02). Retrieved from https://www.tei.org.za/phocadownload/userupload/Citizens-Bribery-Survey-2016-FINAL.pdf

Ethics Institute, 2017. South African Citizens' Bribery Survey, 2017 (Report no. 03). Retrieved from

https://www.tei.org.za/phocadownload/CBS%202017_29%20November%202017_Upload %20to%20web.pdf

Efferson, C. & Vogt, S., 2018. Behavioural homogenization with spillovers in a normative domain. *Proceedings of the Royal Society B: Biological Sciences*, 285(1879).

Fisman, R. & Golden, M., 2017. How to fight corruption. Science, 356(6340), pp.803-804.

- Fisman, R. & Golden, M.A., 2017. *Corruption: What everyone needs to know* 1st ed., New York, NY, USA: Oxford University Press.
- Fjeldstad, O.-H., 2005. Issue: Revenue administration and corruption. *U4 Issue 2:2005*, p.24. Available at: http://www.u4.no/themes/pfm/Revenueissue/revenuefront.cfm.
- Frey, E. & Rogers, T., 2014. Persistence: How Treatment Effects Persist After Interventions Stop
 S. T. Fiske, ed. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), pp.172–179.
- Goffman, E., 1969. *Stigma: Notes on the Management of Spoiled Identity*, New York, NY, USA: Simon & Schuster.
- Heywood, P.M., 2018. Combating corruption in the twenty-first century: New approaches. *Daedalus*, 147(3), pp.83–97.
- Heywood, P.M., 2017. Rethinking corruption: Hocus-pocus, locus and focus. *Slavonic and East European Review*, 95(1), pp.21–48.
- Heywood, P.M., 2014. *Routledge handbook of political corruption* P. M. Heywood, ed., New York, NY, USA: Routledge.
- Hoffmann, L.K. & Patel, R.N., 2017. Collective Action on Corruption in Nigeria: A social norms approach to connecting society and institutions. , pp.1–53. Available at:

https://www.chathamhouse.org/sites/files/chathamhouse/publications/research/2017-05-17corruption-nigeria-hoffmann-patel.pdf.

Jackson, D. & Köbis, N.C., 2018. Anti-corruption through a social norms lens, Bergen, Norway.

- Jamil, T. et al., 2017. Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, 49(2), pp.638–652.
- Jeffreys, H., 1961 *Theory of probability*. (3rd ed.), New York, NY, USA: Oxford University Press.
- Johnsøn, J., Taxell, N. & Zaum, D., 2012. *Mapping evidence gaps in anti-corruption*, Bergen, Norway.
- Kerr, N.L. & Kaufman-Gilliland, C.M., 1997. ".. and besides, I probably couldn't have made a difference anyway": Justification of Social Dilemma Defection via Perceived Self-Inefficacy. *Journal of Experimental Social Psychology*, 33(3), pp.211–230.
- Köbis, N.C. et al., 2016. Prospection in individual and interpersonal corruption dilemmas. *Review of General Psychology*, 20(1), pp.71–85.
- Köbis, N.C. et al., 2015. "Who doesn't?" The impact of descriptive norms on corruption. *PLoS ONE*, 10(6), p.e0131830.
- Köbis, N.C., Iragorri-Carter, D. & Starke, C., 2018. A Social Psychological View on the Social Norms of Corruption. In I. Kubbe & A. Engelbert, eds. *Corruption and Norms - Why Informal Rules Matter*. London, UK: Palgrave Macmillan, pp. 31–52.
- Köbis, N.C., Jackson, D. & Iragorri-Carter, D., 2019. Social norms and corruption Recent trends in empirical research. In P. M. Heywood & A. Mungiu-Pippidi, eds. *The State of the Art in Corruption Research and an Agenda for the Future*.

Köbis, N. C., et al., 2019. Intuitive Honesty Versus Dishonesty: Meta-Analytic Evidence.

Perspectives on Psychological Science, 14(5), 778–796.

- Koelble, T.A., 2018. Caught Between Liberation and Liberalism. In I. Kubbe & A. Engelbert, eds. Corruption and Norms - Why Informal Rules Matter. London, UK: Palgrave Macmillan, pp. 115–132.
- Krupka, E.L. & Weber, R.A., 2013. Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association*, 11(3), pp.495–524.
- Kubbe, I. & Engelbert, A. eds., 2018. Corruption and Norms Why Informal Rules Matter, Palgrave Macmillan.
- van der Linden, S., 2018. The future of behavioral insights: on the importance of socially situated nudges. *Behavioural Public Policy*, 2(2), pp.207–217.
- Loewenstein, G. & Chatter, N., 2017. Putting nudges in perspective. *Behavioural Public Policy*, 1(1), pp.26–53.
- Manning, M., 2009. The effects of subjective norms on behaviour in the theory of planned behaviour: A meta-analysis. *British Journal of Social Psychology*, 48(4), pp.649–705.

Marquette, H. & Peiffer, C., 2015. Corruption and Collective Action, Bergen, Norway.

- Mény, Y., 1996. "Fin de siècle" corruption: Change, crisis and shifting values. *International Social Science Journal*, 48, pp.309–320.
- Mortensen, C. R., et al., 2019. Trending Norms: A Lever for Encouraging Behaviors Performed by the Minority. *Social Psychological and Personality Science*, 10(2), pp. 201–210.
- Mungiu-Pippidi, A., 2017. The time has come for evidence-based anticorruption. *Nature Human Behaviour*, 1(1), p.11.

Nye, J.S., 1967. Corruption and political development: A cost-benefit analysis. The American

Political Science Review, 61(2), pp.417–427.

- OECD, 2018. Behavioural Insights for Public Integrity: Harnessing the Human Factor to Counter Corruption, Paris, France: OECD.
- Oliver, A., 2013. From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy. *Journal of Social Policy*, 42(4), pp.685–700.
- Oliver, A., 2015. Nudging, shoving, and budging: Behavioural economic-informed policy. *Public Administration*, 93(3), pp.700–714.
- Oliver, A. & Ubel, P., 2014. Nudging the obese: a UK–US consideration. *Health Economics, Policy and Law*, 9(3), pp.329–342.
- Olken, B.A., 2009. Corruption perceptions vs. corruption reality. *Journal of Public Economics*, 93(7–8), pp.950–964.
- Pailey, R.N., 2013. Gbagba, New York, NY: One Moore Book.
- Pailey, R.N., 2019. Jaadeh, New York, USA: One Moore Book.
- Paluck, E.L., 2009. What's in a Norm? Sources and Processes of Norm Change. Journal of Personality and Social Psychology, 96(3), pp.594–600.
- Paluck, E.L., Shepherd, H. & Aronow, P.M., 2016. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3), pp.566–571.
- Paprzycka, K., 1999. Normative expectations, intentions and belief. *The Southern Journal of Philosophy*, XXXVII, pp.629–652.
- Peiffer, C. et al., 2019. The surprising case of police bribery reduction in South Africa. *Crime, Law and Social Change*, pp.1–20. Available at: http://link.springer.com/10.1007/s10611-019-09843-8 [Accessed July 22, 2019].

- Peiffer, C. & Walton, G., 2019. Overcoming Collective Action Problems through Anti-corruption Messages,
- Perkins, H.W. & Wechsler, H., 1996. Variation in Perceived College Drinking Norms and its Impact on Alchohl Abuse: A Nationwide Study. *Journal of Drug Issues*, 26(4), pp.961–974.
- Persson, A., Rothstein, B. & Teorell, J., 2012. Why anticorruption reforms fail—Systemic corruption as a collective action problem. *Governance*, 26(3), pp.449–471.
- Pinker, S., Nowak, M.A. & Lee, J.J., 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences of the United States of America*, 105, pp.833–8.
- Rose-Ackerman, S., 2006. *International handbook on the economics of corruption* S. Rose-Ackerman, ed., Cheltenham, UK Northampton, MA, USA: Edward Elgar.
- Rothstein, B., 2011. Anti-corruption: The indirect "big bang" approach. *Review of International Political Economy*, 18, pp.228–250.
- Rothstein, B., 2000. Trust, Social Dilemmas and Collective Memories. *Journal of Theoretical Politics*, 12(4), pp.477–501.
- Rothstein, B. & Varraich, A., 2017. *Making Sense of Corruption*, Cambridge: Cambridge University Press.
- Scharbatke-Church, C. & Chigas, D., 2016. *Facilitation in the Criminal Justice System*, Boston, Massachusetts. Available at: fletcher.tufts.edu/IHS.
- Schram, A., Zheng, J. Di & Zhuravleva, T., 2019. *Contagious corruption : cross-country comparisons*, Florence, Italy.
- Soraperra, I. et al., 2019. *A market for integrity An experiment on corruption in the education sector*, CREED Working Paper, Amsterdam, Netherlands.

Starke, C., Naab, T.K. & Scherer, H., 2016. Free to expose corruption: The impact of media

freedom, internet access, and governmental online service delivery on corruption. *International Journal of Communication*, 10, pp.4702–4722.

- Stephenson, M.C., 2018. Corruption as a self-reinforcing "trap": Implications for reform strategy, QoG Quality of Government Institute Working Paper.
- Stibe, A. & Cugelman, B., 2016. Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In *Lecture Notes in Computer Science*. pp. 65–77. Available at: https://link.springer.com/chapter/10.1007/978-3-319-31510-2_6 [Accessed July 26, 2019].
- Sunstein, C.R., 2015. On Interesting Policymakers. *Perspectives on Psychological Science*, 10(6), pp.764–767.
- Tankard, M. & Paluck, E.L., 2016. Norm as a Vehicle for Social Change. Social Issues and Policy Review, 10(1), pp.181–211.
- Thaler, R. & Sunstein, C., 2008. Nudge: Improving decisions about health, wealth, and happiness, Yale University Press, USA.
- Vogt, S. et al., 2016. Changing cultural attitudes towards female genital cutting. *Nature*, 538(7626), pp.506–509.
- Winters, M.S., Testa, P. & Fredrickson, M.M., 2012. Using Field Experiments to Understand Information as an Antidote to Corruption. In D. Serra & L. Wantchekon, eds. *New Advances in Experimental Research on Corruption*. Emerald Group Publishing Limited, pp. 213–246.
- Zhao, H., Zhang, H. & Xu, Y., 2017. Effects of perceived descriptive norms on corrupt intention: The mediating role of moral disengagement. *International Journal of Psychology*.