



Peiffer, C. (2021). The Curse of Good Intentions: Why Anticorruption Messaging Can Encourage Bribery. *American Political Science Review*. <https://doi.org/10.1017/S0003055421001398>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1017/S0003055421001398](https://doi.org/10.1017/S0003055421001398)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via CUP at <https://doi.org/10.1017/S0003055421001398>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

The Curse of Good Intentions: Why Anticorruption Messaging Can Encourage Bribery

NIC CHEESEMAN *University of Birmingham, United Kingdom*

CARYN PEIFFER *University of Bristol, United Kingdom*

Awareness-raising messages feature prominently in most anticorruption strategies. Yet, there has been limited systematic research into their efficacy. There is growing concern that anticorruption awareness-raising efforts may be backfiring; instead of encouraging citizens to resist corruption, they may be nudging them to “go with the corrupt grain.” This study offers a first test of the effect of anticorruption messaging on ordinary people’s behavior. A household-level field experiment, conducted with a representative sample in Lagos, Nigeria, is used to test whether exposure to five different messages about (anti)corruption influence the outcome of a “bribery game.” We find that exposure to anticorruption messages largely fails to discourage the decision to bribe, and in some cases it makes individuals more willing to pay a bribe. Importantly, we also find that the effect of anticorruption messaging is conditioned by an individual’s preexisting perceptions regarding the prevalence of corruption.


Corruption has long been recognized as a major threat to sustainable development, both because it leads to the waste of public resources and because it can distort incentives for officials and citizens alike. However, efforts to combat corruption by encouraging political leaders and bureaucrats to implement the rule of law have met with limited success (Johnson, Taxell, and Zaum 2012; Rocha Menocal and Taxell 2015). Anticorruption programs often fail to recognize the incentives of influential individuals, who—if they are operating in systemically corrupt environments characterized by weak institutions—have little reason to enact meaningful reform (Khan, Andreoni, and Roy 2019). Partly for this reason, there has been a broadening of emphasis toward influencing public opinion. This makes intuitive sense. If bringing about sustainable anticorruption reform requires changing the incentives facing political leaders, increasing popular demand for “clean” government is one way to make supporting corrupt practices more costly.

Donors, nongovernmental organizations, and anticorruption agencies have invested millions of dollars in encouraging ordinary citizens to condemn graft. Yet there has been limited research to test the effect of these efforts. To date, only five studies have examined the effect of anticorruption messaging (Corbacho et al. 2016; Kobis et al. 2019; Peiffer 2017; 2018; Peiffer and Walton 2019), and only two of those examined whether messaging influences an individual’s willingness to engage in corruption. In both studies, exposure to a

message failed to encourage citizens to reject bribery (Corbacho et al. 2016; Kobis et al. 2019). These limited findings suggest that ordinary citizens may not respond to anticorruption messages in the desired way. We need to know much more about why this happens and the implications for the fight against corruption. This paper contributes to and advances this discussion by exploring the effects of anticorruption messaging on bribery in Lagos, Nigeria.

The logic underpinning existing messaging efforts is intuitive. A nearly universally accepted adage in anticorruption work is that because graft “lives in the shadows,” a critical step to fighting corruption is to illuminate and bring popular awareness to it. The assumption behind raising awareness is that doing so will strengthen demands for clean government. This seemingly uncontroversial idea has served as the starting point for many anticorruption policies and is echoed in the call made in the 2004 *United Nations Convention Against Corruption* (UNCAC), which asks governments to invest in raising the public’s awareness of the “existence, causes and gravity of and the threat posed by corruption” (United Nations 2004, 15). There is now a nearly ubiquitous push among donors and nongovernmental organizations to raise public awareness through billboards, posters, murals, radio and television shows, and so on.

Research into the effect of these messages is at an early stage but has already raised concerns. A small number of qualitative studies have suggested that awareness-raising campaigns have a limited effect, either because most individuals do not see the messages, or because they are not influential enough to outweigh other kinds of information (Baez Camargo 2017; Hoffman and Patel 2017). A recent strand of quantitative research also suggests that by making individuals aware of pervasiveness of corruption, these messages may make citizens more despondent and undermine their willingness to resist corruption (Peiffer 2017; 2018). In other words, telling people

Nic Cheeseman , Professor of Democracy, International Development Department, University of Birmingham, United Kingdom, n.cheeseman@bham.ac.uk.

Caryn Peiffer , Senior Lecturer in International Public Policy and Governance, School for Policy Studies, University of Bristol, United Kingdom, Caryn.peiffer@bristol.ac.uk.

Received: March 12, 2021; revised: August 11, 2021; accepted: November 12, 2021.

about corruption may make the fight for a “clean” government seem helpless. One study argues that this is not only the case for messages that explicitly stress how pervasive corruption is: even much more careful, targeted and up-beat statements that emphasize anticorruption efforts may have this effect (Peiffer 2018). This finding is of profound importance because it implies that any campaign that primes the public to think about corruption may do more harm than good.

In this paper, we move the literature forward in five main ways. First, we focus on the influence of messaging on behavior, as opposed to self-reported attitudes or dispositions toward corruption. Significantly, this is the first study to test the effect of anticorruption campaigning on behavior by showing messages to individuals and playing a “bribery game” in which they could win real money depending on whether they were willing to pay a bribe. It therefore offers a more realistic test of the effect of anticorruption messages than previous studies.

Second, we test the effect of five different messages—more than any previous study—that represent a broader range of framings, tones, and themes. Third, our study looks at a new case—Lagos, Nigeria—which enables us to say more about the generalizability of the emerging critique of anticorruption messaging. Fourth, we add nuance to the findings of the existing literature by demonstrating that anticorruption messages have different effects on different types of individuals, which is important for understanding when messaging is most likely to generate problematic unintended consequences. Finally, our analysis assesses not only how preexisting beliefs about the prevalence of graft shape the influence of anticorruption messaging but also the factors that are associated with these beliefs.

Our findings address the remarkable resilience of corruption in many developing countries, but they also offer a glimmer of hope. In line with previous research, we find that exposure to anticorruption messages fails to discourage corrupt behavior and in some cases *makes individuals more willing to bribe*. However, this effect is not universal. Instead, the influence of anticorruption campaigns is conditioned by an individual’s preexisting perceptions regarding the prevalence of corruption. Among those who already believed that corruption is widespread, anticorruption messages typically backfired, encouraging corrupt behavior in our game. Tellingly, being pessimistic about corruption is not randomly distributed; instead, it is more common among those who have personally witnessed corruption, as one would expect. Among this group, which we refer to as “pessimistic perceivers,” the paper documents a vicious cycle: witnessing corruption is associated with pessimism about the pervasiveness of corruption within society, and when a message primes the issue of corruption, pessimists are more likely to be willing to engage in bribery than nonpessimists.

The prospects for reform appear bleak if we focus only on “pessimistic perceivers,” but the situation is more positive among those who do not see corruption to be such a big issue. Among these individuals, most messages have no influence while one—the message that stresses the direct effect of corruption on

individuals’ through the lens of taxation—made respondents significantly less likely to pay a bribe in our game. This implies that although anticorruption messaging may not have a positive effect overall, certain specifically targeted messages have the potential to strengthen public resolve to resist corruption.

That the effect of messages is shaped by perceptions of the pervasiveness of corruption has important implications for both academic and policy debates. Empirically, our findings demonstrate the need to differentiate between types of citizens and the limitations of assuming that messages have a standard effect. Theoretically, our findings suggest that the main reason why anticorruption messages have negative consequences is that they reinforce perceptions that corruption is widespread and thus reinforce collective action problems. In turn, this highlights the power of priming (Krosnick and Kinder 1990; Van Duyn and Collier 2019) and the danger that government media campaigns may have unintended negative effects on popular attitudes. The implications for public policy and the design of anticorruption interventions are profound: untargeted messaging is not just a waste of money; it may actually be making it harder for other strategies to succeed.

LITERATURE REVIEW AND HYPOTHESES

Anticorruption awareness-raising efforts are underpinned by the assumption that when the public becomes more aware of graft it also becomes less tolerant and willing to engage in it. The idea that messaging can shape behavior resonates well with broad findings from the field of experimental political psychology, which has shown that messaging can significantly shape political behavior (Berinsky et al. 2011; Carter, Ferguson, and Hassin 2011; Nosek, Graham, and Hawkins 2010). Significantly, it has even been established that political messages can shape behavior even when they are not seen to be particularly influential by those exposed to them (Erisen, Lodge, and Taber 2014).

Policy documents on awareness-raising such as the UNCAC rarely specify how messages are expected to reduce corruption, however. This is important, because a growing literature has cast doubt on the efficacy of anticorruption messages, suggesting that such campaigns may be “priming” the issue of widespread corruption, with potentially negative consequences. Through “priming,” messages change how a person behaves by making them think more about the issue than they would have done otherwise (Brody and Page 1972; Lenz 2009; Riker 1986). This has led to considerable concern that information campaigns that draw attention to the issue of pervasive corruption will have the opposite effect to that intended: encouraging corruption rather than its condemnation.

Priming Widespread Corruption: The Backfire Effect

The concern that anticorruption messaging is counterproductive finds its roots in literature that casts

systemic corruption as a collective-action problem. This research suggests that when a person believes that most people engage in corruption, they will also be less likely to believe they should abstain from it (Mungiu-Pippidi 2011; Persson, Rothstein, and Teorell 2013; 2019; Rothstein 2011). Instead, they may believe that if they do not engage in corruption they will be sanctioned socially for deviating from the perceived norm; punished materially by having to pay more, wait longer, or go without a service or good; miss out on an opportunity to benefit from a corrupt exchange; and/or not be able to make a difference in corruption patterns by abstaining.

There is also growing evidence that individuals are more likely to engage in certain behaviors if they believe that others have already done so (Manning 2009; Tankard and Paluck 2016). Research on social norm “nudging,” for example, has found that behavioral change sometimes occurs when individuals realize that their peers have already started to behave in a different way (Bicchieri and Dimant 2019). Significantly, recent research conducted in Nigeria has come to similar conclusions. As Hoffman and Patel (2017) write, many Nigerians have overly pessimistic beliefs about the fight against corruption, in part because they believe that fellow citizens are more supportive of corrupt activities than is really the case.

These overlapping literatures suggest that raising awareness of corruption among those who already think it is widespread may be particularly problematic. Making what we call pessimistic perceivers—that is, those who already believe corruption is pervasive—more aware of the problem may encourage “even former noncorrupt actors to take part in the corrupt game” (Persson, Rothstein, and Teorell 2013, 464–5). In turn, this implies that the influence of anticorruption awareness-raising on corrupt behavior will vary depending upon preexisting beliefs about corruption. For pessimistic perceivers, any messaging at all—even an inspiring message about anticorruption—may encourage corrupt behavior. For others who are less convinced that corruption is endemic and so do not have the same social beliefs to prime, a message about corruption may not backfire.

To the best of our knowledge, only five studies have so far assessed the influence of anticorruption messages (Corbacho et al. 2016; Kobis et al. 2019; Peiffer 2017; 2018; Peiffer and Walton 2019).¹ Although none of these projects scrutinized whether pessimistic perceivers respond differently to messaging, the findings of two of these studies are especially relevant. Using a household survey of 1,000 respondents in Jakarta, Indonesia, Peiffer (2017; 2018) examined what influence four messages about corruption and anticorruption had on attitudes and perceptions. Significantly, all four messages were found to elicit *the same degree of*

increased worry about the harms that corruption causes, decreased pride in the government’s efforts to fight corruption, and decreased confidence that ordinary people can participate easily in civic anticorruption activities (Peiffer 2018). Peiffer (2017) also found that these messages had a negative influence on willingness to join an anticorruption organization or protest and little influence on willingness to report corruption. This is surprising because the messages differed in their framing, tones, and contents, and one message was explicitly designed to mobilize citizens in just this way. Though Peiffer (2017; 2018) did not examine what effect messaging has on behavior, which is our focus, the conclusions drawn from her studies lend compelling support to the notion that any message at all about (anti)corruption could backfire.

Two of the other existing studies reinforce this point, although as they only examined the effect of one message they tell us less about whether different messages elicit a universal reaction. Corbacho et al. (2016) conducted a survey experiment with more than 4,000 respondents in Costa Rica in 2013. They found that respondents exposed to a message that a growing number of Costa Ricans were practicing corruption were more likely to say they would be willing to bribe a police officer. A more recent experiment implemented by Kobis et al. (2019) in South Africa also found that anticorruption messaging did not influence willingness to bribe. Although this study did not find that putting up anticorruption messages in the town of Manguzi has a counterproductive effect, its findings inspire little confidence in existing strategies.

In this paper, we explicitly test for the interaction implied by these collective-action theory and social norms literatures, hypothesizing that

H1: Exposing those who believe that corruption is widely practiced to any message about corruption will increase willingness to pay a bribe.

Influencing Behavior through Positive Persuasion

Although a range of messages may influence behavior similarly by *priming* the issue of corruption as predicted in H1, it is also possible that exposure to different messages may lead to different consequences. This is demonstrated by the fifth anticorruption messaging study, Peiffer and Walton’s (2019) survey experiment in Port Moresby, Papua New Guinea. Using a sample of over 1,500 respondents, their research examined the influence of four messages on willingness to report corruption. Exposure to three of the messages that emphasized the rule of law, criticism of corruption by religious leaders, and the pervasiveness of corruption was found to have no significant effect on attitudes toward reporting corruption. However, exposure to a fourth message that stressed that corruption is a “local” issue encouraged favorable attitudes about reporting corruption. This stands out as an exception in the literature—an anticorruption message that was found

¹ Several scholars have researched how messages about corruption influence political behavior. For an excellent meta-analysis, see Incerti (2020). Most examine how messages about corruption influence voting (see Anduiza, Gallego, and Munoz 2013; Chong et al. 2015; Winters and Weitz-Shapiro 2013).

to have an “intended” effect—and suggests that some messages may be more productive than others.

One way to make sense of Peiffer and Walton’s (2019) finding is that messaging may influence behavior through *persuasion*. In contrast to priming, which makes a person think more about an issue, a persuasive message changes the way a person thinks about an issue and may therefore prompt different actions (Brody and Page 1972; Lenz 2009; Riker 1986). Numerous studies on message framing, for example, show that how an issue is presented can significantly change how the public thinks about it (see Druckman 2001; Jacoby 2000; Sniderman and Theriault 2004). Given Peiffer and Walton’s (2019) finding that a message with a “local” framing encouraged willingness to report corruption, we might expect that messages that emphasize the direct effect of graft on the community, or even the individual, will discourage corruption.

To an extent, this reflects existing practice. “Community-based anticorruption programmes,” which are “physically and conceptually located in a community to fight and counter corruption,” have been found to be fairly effective (Richards 2006, 5). Reinikka and Svenson (2011), for example, have demonstrated that informing citizens about localized education funding in Uganda inspired greater citizen monitoring that ensured that money flowed to schools as intended.

At the individual level, corruption experts such as Kenya’s former anticorruption tsar John Githongo have long argued that one reason why corruption did not always rank as a high priority for citizens in Africa is that individuals do not always see the connection between elite-level graft and their own personal circumstances (Cheeseman and Klaas 2018, 84–8). Indeed, early Afrobarometer surveys found that only 1–5% of respondents saw corruption as the most important issue facing their countries (Mattes et al. 2000). One plausible explanation is that citizens felt that corruption was a marginal matter relative to, say, education and law and order because the connection between the theft of public funds and the limited resources available for these services was partly hidden to them. Some individuals may even endorse corruption—or not see graft as corruption at all—in cases where they feel that clientelism and patronage benefit their own group (Cheeseman, Lynch, and Willis 2020). As Olivier de Sardan (1999) has influentially argued, legitimizing corruption in this way is considerably easier in countries where graft is seen as a basic fact of life. Significantly, qualitative research on Nigeria has often emphasized exactly this point (e.g., Smith 2008).

The literature on social contracts in Africa has come to a similar conclusion from a different perspective. It is common for researchers to argue that one reason that there are not greater demands for political accountability in Africa is that relatively few citizens pay direct (i.e., income) tax (Moore, Prichard, and Fjeldstad 2018; Prichard 2015). The assumption here is that because many individuals are too poor to pay income tax, or the state is too weak to collect it, a large part of the population does not see the connection between their money and government revenue. Indirect taxes such as

sales taxes or value added tax (VAT) do not have the same effect precisely because they are rarely explicitly stated and may not be paid by those who rely on informal markets. Most notably, recent research—including a number of studies in Nigeria (Bodea and LeBas 2016)—has found that taxation plays an important role in the formation of social contracts and that tax payment generates stronger demand for accountability (Cheeseman and de Gramont 2017). This is particularly relevant in Lagos, where the state government has, for over a decade, focused on expanding and extending the “tax net,” including to cover those in lower income groups (Cheeseman and de Gramont 2017).

These diverse literatures therefore reach the same conclusion: individuals will be more willing to condemn corruption if they see an explicit relationship between graft and their own personal finances or locality. This gives rise to an alternative hypothesis to H1:

H2: Messages that emphasize the way in which corruption affects a specific individual and/or their community will have a greater effect in terms of reducing bribe-paying behavior.

Leadership Framing

The efficacy of awareness-raising efforts may also be shaped by other ways of framing messages. Research has demonstrated that citizens often rely on “information shortcuts,” or heuristics, and so messages may be particularly persuasive if they are endorsed by leaders (Kahneman, Slovic, and Tversky 1982; Lupia 1994). The logic here is that when individuals are presented with a message, they may be persuaded of its argument by virtue of the perceived characteristics of the leaders who endorse it (Dewan, Humphreys, and Rubeson 2014).

A more pessimistic, but related, argument has been made about the influence of leadership on grassroots corruption. The often cited “fish rots from the head down” metaphor describes the intuition that, when deciding whether or not to engage in corruption, citizens take cues from public officials (see Rose-Ackerman 2015; Rothstein 2013; Rothstein and Eek 2009). In Rothstein’s (2013, 1021) words, “the ethics of public officials become central here, not only with respect to how they do their jobs but also to the signals they send to citizens about what kind of ‘game’ is being played in the society.”

When applied to anticorruption messaging, the logical extension of the fish-rots-from-the-head-down hypothesis can be bleak. If messaging persuades individuals to think that leaders are in on the corrupt game, citizens may become more willing to participate in corruption. However, the same logic can be applied to generate a more hopeful prediction. If a message can persuade citizens to think that leaders are earnestly fighting corruption, citizens may become more willing to resist corruption themselves. To this end, Peiffer and Alvarez (2016) found that citizens who perceived that the government was effective in

fighting corruption were more willing to personally engage in anticorruption civic activities, such as protesting. Persuaded by such messaging, citizens may feel encouraged that meaningful reform is possible—as appears to have been the case in Lagos, at least up until 2015 (Cheeseman and de Gramont 2017). From the optimistic framing of this logic, we derive a third hypothesis:

H3: Awareness-raising messages that emphasize that leaders are actively fighting corruption will discourage a willingness to engage in corruption.

In addition to testing the hypotheses set out above, our research contributes to the emerging literature on awareness-raising in five ways. First, we focus on corrupt behavior, as opposed to self-reported willingness to bribe (Corbacho et al. 2016), perceptions of corruption (Peiffer 2018), or willingness to report corruption (Peiffer and Walton 2019). We do this by building on the work of Kobis et al. (2019), playing a bribery game that gives individuals an opportunity to win real-world money. Second, we test for the effects of multiple messages. Third, our sample is large and representative of Lagos State (see Appendix A), and so our findings represent a more reliable assessment of the effect of corruption messaging than studies that have relied on smaller and/or nonrandom samples (Kobis et al. 2019; Peiffer and Walton 2019). Fourth, our analysis tests for whether the effect of a message is dependent upon prior perceptions of how common corruption is for the first time. Finally, this enables us to make a fifth contribution: we identify who is most likely to have pessimistic attitudes about corruption.

METHODOLOGY

This study uses data from an experiment conducted in Lagos, Nigeria. Lagos is one of world's fastest growing megacities and is run by a directly elected governor as one of the 36 states that make up Nigeria's federal political system (Cheeseman 2015). Lagos is also the former political capital of Nigeria and is widely recognized as the country's economic capital.

Four features of Lagos make it particularly suitable for this study. First, corruption is thought to be a considerable problem there. Of the 183 countries assessed in Transparency International's 2019 Corruption Perception Index, Nigeria is ranked 146th for its control of corruption, which is on par with Angola and Bangladesh (Transparency International 2020). Second, corruption is not a socially taboo topic (Smith 2008), and so recruitment of participants was not problematic. Third, as an ethnically and religiously diverse context with high levels of both poverty and inequality, Lagos shares many of the features that are often said to complicate development efforts. Finally, despite this challenging context, the Lagos State Government has had some public "wins" in its efforts to increase taxation and provide better services

to citizens, massively expanding the tax take (Bodea and Lebas 2016).² This makes Lagos a good location to test messages about both taxes and the role of political leaders.

The study ran in Lagos from December 21, 2019 to January 12, 2020. We followed Afrobarometer's established protocol (Afrobarometer 2020) to recruit a representative sample of individuals at the household level.³ For the overall survey, 2,572 individuals took part, of which 1,200 participants played our bribery game.⁴

Instrument

The bribery game discussed in this paper was one element of a larger project that included a broader instrument that asked participants a series of survey questions. With all participants, enumerators started by reading a short introductory paragraph that described the study's purpose as wanting to "learn what citizens think about public services and the experiences they have with public officials" (Appendix A). Enumerators explained that responses would be treated confidentially and that participants could withdraw at any time. Study participants were then randomly assigned to one of six groups: *control*, *widespread*, *religious*, *government success*, *local fight*, or *taxes* ($n = 400$ in each).

If assigned to the *widespread*, *religious*, *government success*, *local fight*, or *taxes* group, questions about basic personal information (age, gender, education) were followed by the respective anticorruption treatments (messages). After exposure to the treatment (or not for the control group, which proceeded directly to the next set of questions), participants were asked survey questions to gauge their perceptions of corruption and anticorruption efforts.

After these questions, half of the participants were randomly selected to play a game in which they would be able to win a small amount of money. The participants of the game were evenly distributed across the treatment groups ($n = 200$ per group). This paper focuses on the influence of the treatments on the decision to pay a bribe in the game.

² Although the legacy of Governor Babatunde Fashola (2007–2015), whom *The Economist* magazine once described as "a rare good man" (May 5, 2011), was subsequently called into question by allegations that he misused public funds while in office.

³ Information about the recruitment process and the demographic characteristics of the sample are provided in Appendix A. A description of ethical considerations and procedures of the research are found in Appendix J.

⁴ The project could not afford for all 2,400 to play the game. A total of 1,372 participants were approached to play, of which 1,200 decided to play. The percentages of game players among total participants approached to play, by treatment group, are 86% (*control*), 87% (*widespread*), 85% (*local*), 91% (*tax*), 89% (*government success*), and 88% (*religious*). The decision to play is not significantly associated with treatment group assignment and is therefore treated as independent of any influence of the treatments tested. For replication data and do file for all analyses, see Cheeseman and Peiffer (2021).

Treatments

Each of the treatments was a paragraph long (see Appendix C).⁵ The *widespread* message sought to make the widespread nature of corruption in Nigeria salient. This treatment was tested because awareness-raising efforts regularly publicize how widespread and harmful corruption is (Jones 2011).

By describing the outcry against corruption from many different religious leaders, the *religious* treatment provides a message to test H3. Nigeria is a highly religious country, and so an anticorruption message endorsed by various religious leaders may be especially persuasive. The treatment used is indicative of anticorruption awareness-raising efforts in several countries, with one example being a very recent campaign involving several religious leaders in Uganda, which was supported by international donors (Uganda Media Centre 2020). The logic underpinning these efforts is that religious leaders likely have a perceived higher moral authority than public-sector actors, and so carry more influence (Marquette 2012).

The *government success* treatment mentioned salient achievements of the Lagos State Government in fighting and reducing corruption and especially emphasized the leadership of Governor Fashola. This message allows for an additional test of H3. By emphasizing the strides made by Fashola's government, the message seeks to persuade others to follow suit and resist corruption. The expectation of a positive outcome from a government effectiveness message addresses Peiffer and Alvarez's (2016) finding that perceived government effectiveness in anticorruption is positively associated with a greater willingness to report corrupt behavior. In practice, anticorruption agencies often publicize their wins in the fight against corruption and do so through messaging campaigns.

The *local fight* treatment frames corruption as an issue that acutely influences local communities and argues that the fight against corruption should prioritize local, communal efforts. It therefore allows us to test H2. The wording was inspired by the treatment used in Peiffer and Walton's (2019) study in Papua New Guinea, where exposure to a similar message encouraged positive attitudes toward reporting corruption. Like the Papua New Guinea context, many Nigerians identify strongly with their local, often ethnic, communities.

Finally, the *taxes* treatment allows for an additional test of H2, as it suggests that corruption represents the theft of taxes and fees that ordinary citizens pay on a daily basis. Lagos represents a good location to test this hypothesis given that the efforts of the previous two state governments to expand the tax base have been high profile, sensitizing citizens to the relationship between paying taxes and the provision of public services.

⁵ A pilot study was conducted to ensure that the messages were well understood.

Our treatments were designed to mimic the kinds of information that are provided to participants in anticorruption roadshows and training sessions. It is worth noting, however, that unlike some public relations campaigns, our messages do not use heavily emotive language, like "corruption kills"—a phrase sometimes used on anticorruption billboards. Moreover, as our treatments were only provided in text form, participants were not exposed to associated provocative imagery, another common feature of awareness-raising. Our treatments were therefore not as "strong" as some—but not all—messaging used in practice and therefore present a stricter test of our hypotheses; if our "weaker" treatments are found to be effective, then stronger messaging may have a more powerful effect.

Bribery Game

The bribery game tests whether an individual's willingness to bribe is influenced by the treatment they received. To do this, participants played a version of a dictator game, in which the participant knows that paying a bribe can improve the payout to themselves but will also reduce the payout to another individual. We chose this format because in real life when an individual pays a bribe to access a service or skip a queue (for example at a hospital) they do so at the expense of someone else (for example another patient who must wait longer to be seen). As in our bribery game, the bribe payer most likely does not know who is affected but will be aware that paying the bribe may have an effect on somebody else. The format of our game therefore represents a good approximation of situations that ordinary citizens regularly experience.

More specifically, at the start of the game each participant was told that they were playing with an anonymous second player elsewhere in the city and that the game was overseen by a third party. To start the game, the participant was told that both they and the other player had been given \$5 to play with and that they had only one decision to make: whether to offer the third party a bribe of \$1.50. The participant was told that the second player could not offer a bribe and that if the participant decided to bribe they would automatically be given the \$5 that had initially been intended for the other player. Consequently, the participant would leave with \$8.50, whereas the other player would leave the game with nothing. By contrast, the participant was told that if they chose not to pay a bribe, both they and the other player would walk away with the \$5 they were both originally allocated. The choice to use the term "bribe" in our game was explicit so that there was no ambiguity for our participants on what the game was designed to approximate. And, as our participants stood to make real money, the game represents a test of how individuals behave when their material interests are at stake.

The game was played on a tablet, with participants entering their decisions without enumerators being able to see whether they had paid a bribe. Participants were told that there was no correct or incorrect way to play the game and that it was for each participant to

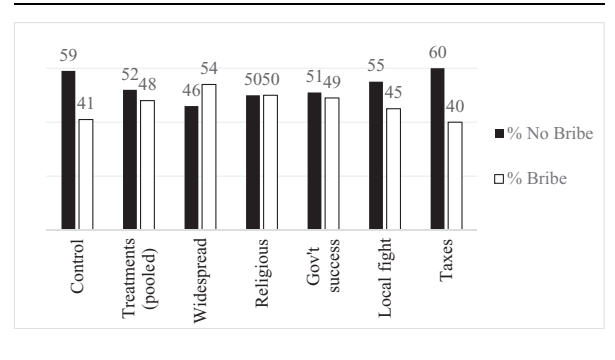
decide for themselves. In the interest of safety and so that enumerators did not have to handle money, all payments from the game were made via mobile phone transfer. At the end of the game, the participant was debriefed and told that in reality there had not been any other players—that is, there was no second player and the third party was simply the tablet, which responded to their decision by calculating the correct payouts.

Although our game represents a significant advance on previous studies by being played in respondent's own homes and with real money, it remains a simulation and there are two important ways in which it diverges from real-life situations. Most obviously, there is no uncertainty in our game, for example about how much money to offer as a bribe, whether it will be accepted, and whether the attempt to pay a bribe will be reported to some higher authority. By contrast, uncertainty is a component of corruption in some cases—for example, when the agent of a multinational company is not sure whether offering a bribe to a minister will ensure they are awarded a lucrative contract. As a result, our game is a better reflection of the situation that individuals find themselves in when there is a widespread understanding of how much a bribe should be in a given context. This is a significant caveat, but it is important to note that this is the type of corruption that is experienced the most by ordinary citizens, who are the focus of our research. The request to pay “something small” in order to be left alone by traffic police or to access a public service, for example, is a common feature of life in systemically corrupt contexts. In these types of exchange, there is often no question of whether the official will accept a bribe and the amount is generally well-established. With its non-negotiable bribe amount, our game reflects bribery that many people encounter.

The second deviation from ordinary life is that there is not an actual public service provided in our game—no participant stands to gain/lose access to actual health care or education. This was a necessary feature of our game, as it was neither logistically feasible nor ethically acceptable to implement a game in which access to services would have been at stake. Therefore, it is important to be careful about how we interpret our findings and apply them to the real world. It may be, for example, that individuals are less willing to pay a bribe when they know that they are denying someone else a concrete service rather than simply a financial return.

There is good reason, however, to think that our game has external validity. The bribe amount of US \$1.50 was chosen because it is close to the amount that Nigerians are most likely to have to pay—around US \$0.10–US\$0.50 at police roadblocks—and falls in the middle of the modal category of bribes, which, according to a survey conducted in 2019, is US\$3.63 and under (UNODC Research 2019). The total that the player can win, \$8.50, is a meaningful amount in a country where the minimum wage is \$79 a month, or around \$2.10 a day (BBC 2019). Furthermore, payment was framed and paid in terms of money, as this is how 93% of all bribes in Nigeria are transacted (UNODC Research 2019, 6). We also deliberately did not play the game in a

FIGURE 1. Distribution of Decision to Bribe in Game, by Treatment Group



lab so that individuals would not be operating in an artificial environment.

Moreover, there is strong evidence that participants believed that they were taking part in a game in which other people would be disadvantaged. Had participants believed that they were simply playing a simulation with no other participants, we would have expected every participant to pay a bribe in order to maximize their return. Instead, a minority of participants paid a bribe in our game (Figure 1), which suggests that most participants chose not to because they believed that it would negatively affect another real person. The bribery rate among participants in the game (41% for the control group, 48% for the treatment groups, pooled) is close to the estimated bribery rate for Nigeria as a whole, which stands at 44% (Pring and Vrushi 2019).⁶

Indeed, it is likely that our approach underestimates the tendency toward bribe payment because the participant is made explicitly aware that any decision to pay a bribe will have a negative and known effect on another person. This is rather different to the situation in which someone is asked for a bribe from a police officer, where it is less clear that anyone will lose out if the bribe is paid. Where bribe paying is routine, an individual may not enter into an explicit moral calculation, engaging in no conscious deliberation regarding who wins and who loses. Thus by forcing participants to focus on the negative financial effect that paying a bribe will have on someone else, our game both raises the moral stakes and makes them more explicit. While recognizing the limitations of any simulation and the need to interpret our results carefully—and in conversation with prior work that has studied the effect of anticorruption messages on popular perceptions—it is therefore also important to stress that if anything our findings are likely to *underestimate* popular willingness to engage in corruption.

Estimation Strategy

Pairwise difference-in-means (DIM) tests are appropriate for evaluating the influence of messaging in an experiment like this, when an assumption can be made

⁶ I.e., the percentage of people who paid a bribe in the last 12 months.

that the only difference between respondent groups is that they received different treatments or did not receive a treatment at all (the control group). The DIM tests were run on basic demographic indicators, and the results revealed that the mean level of poverty in the *local* and *control* groups was significantly higher than that of the *religious* group (Appendix D). There were no significant differences among the six groups (five treatment groups and control group) with respect to the rest of the demographic data collected (gender, age, and education).⁷ Therefore, instead of using DIM tests, logistic regressions were conducted to determine how exposure to the treatment messages influenced bribery payment in the game. These analyses allow us to control for the potential influence of poverty so that we ensure that varying reactions to treatments across groups are not due to varying poverty levels. Logistic analyses are also appropriate to use given that the decision to pay a bribe in the game is binary.⁸

RESULTS

We ran two logistic regressions to test our three hypotheses. In the first logistic regression we examine what effect exposure to the five messages had on bribery payment in our game. This allows us to directly test H2 and H3, where the collective expectations are that exposure to the *religious*, *government success*, *local fight*, and *taxes* messages would discourage bribery. H1, on the other hand, requires us to test for the influence of an interaction term (between perceptions of widespread corruption and exposure to messaging), which we do in our second analysis.

The five messages we test are substantially different from one another, both in spirit and in structure. This design is unlike some messaging experiments that examine the influence of slightly different messages, with the goal of comparing how subtle differences between messages influence a scrutinized outcome. In contrast, our objective was to test what effect exposure to each of the five substantially different messages has on bribery. Due to our messages' sufficient differences, we focus our analysis on comparing the influence of exposure to each message to the behavior of the control group rather than on comparisons between treatment groups. Therefore, in both analyses, the baseline group is the control group and all reported messaging effects articulate comparisons between those who were exposed to a message and those who were not.

As standard logistic regression coefficients are difficult to interpret, we also report predicted probability shifts and focus on them in our interpretation of the results. Predicted probability shifts show the direction of association between exposure to the messages and

the decision to pay a bribe in our game—a positive shift illustrates that exposure to a message increased the likelihood that an individual would decide to pay a bribe. Predicted probability shifts also indicate the substantive size of the effect that exposure to a message has on the decision to pay a bribe in our game. They specifically articulate the estimated extent to which exposure to each message changed the probability of paying a bribe.⁹

Do Leadership, Community, or Tax Framings Discourage Bribery?

The results in Table 1 do not provide support for H2 and H3. In fact, with respect to H3, the results show that instead of persuading people to reject the opportunity to pay a bribe, exposure to the two messages that emphasize the role of leaders in the fight against corruption—*religious* and *government success*—had the opposite expected effect, *encouraging* bribery in our game. Specifically, exposure to the *religious* message is significantly and positively associated with choosing to pay a bribe, using a significance threshold of $p < 0.10$. The estimated predicted probability shift shows that exposure to the *religious* message significantly increases the predicted probability of paying a bribe in the game by a sizeable 10 percentage points. Although exposure to the *government success* treatment is not statistically significant (at $p < 0.10$), it is notable that it is close to significant ($p = 0.114$) and its estimated positive association suggests that exposure

TABLE 1. Effect of Messaging on Bribery Payment

	<i>b</i>	SE	<i>p</i>	$\Delta p.p.$
<i>Treatment groups</i>				
Widespread	0.51	0.204	0.013	0.13
Religious	0.41	0.203	0.046	0.10
Gov't success	0.32	0.204	0.114	0.08
Local fight	0.14	0.204	0.493	0.03
Taxes	-0.06	0.205	0.773	-0.02
<i>Control</i>				
Poverty	0.24	0.070	0.001	
Constant	-0.64	0.167	0.000	
<i>N</i>	1,188			
Pseudo <i>R</i> ²	0.01			
Likelihood ratio	23.27			
Prob > χ^2	0.000			

Note: Displayed across the columns are coefficients (*b*), standard errors (SE), *p* values, and predicted probability shifts ($\Delta p.p.$).

⁷ Details of how demographic variables were measured and the results of the DIMs appear in Appendix B and Appendix D.

⁸ Our analyses were conducted on the whole sample that played the game (ITT) rather than just of those who paid attention to the messages (compiler effect).

⁹ These shifts were calculated from postestimated analyses using marginal effects in Stata, with the values of other variables in models held constant.

to this message may also tend toward encouraging bribery.

The results also do not support H2. Instead of discouraging bribery in the game, the results in Table 1 show that there is no statistically significant difference in the proclivity to bribe between those who were exposed to the *local fight* and *taxes* messages and those who were not shown any message (the baseline *control* group).

Exposure to the *widespread* message, however, is significantly and positively associated with bribery. Compared with those who were not shown a message, those in the *widespread* treatment group had a 13-percentage-point greater predicted probability of paying a bribe in our game. This finding echoes the warnings of other studies that anticorruption messaging can backfire (Corbacho et al. 2016; Peiffer 2017; 2018; Persson, Rothstein, and Teorell 2013). It also provides circumstantial evidence to support the idea that messages about widespread and pervasive corruption may encourage so-called corruption fatigue, whereby people who are already overwhelmed with the problem of widespread corruption resign themselves to “accept reality” and go with the corrupt grain (Peiffer and Alvarez 2016).

As they relate to the estimated influence of exposure to messaging, the results in Table 1 are robust to an alternative modeling specification. Because the messages tested are substantially different from each other, we also ran five additional logistic regressions—one for each treatment group—to test whether examining exposure to each message, in isolation, is estimated to have influenced the decision to bribe in our game to the same extent (see Appendix H). The estimated effect of exposure to each message in those five logistic regressions is nearly identical to what is reported in Table 1.

Finally, given that our poverty measure is positively and significantly associated with the decision to bribe in Table 1, it is important to reiterate that we have controlled for poverty in this analysis because the mean level of poverty in the *local* and *control* group were significantly higher than in the *religious* group. Our results can therefore not be driven by this variable,¹⁰ but it is nonetheless worth noting that poverty has an intuitive effect. In four of the five regressions presented here, poverty increased the willingness of individuals to pay a bribe. It is tempting to simply assume that this relationship exists because poorer respondents have a greater need for the higher payout that can be earned in our game by offering a bribe. It may also, however, be related to the fact that poorer citizens are more reliant on state services and are therefore more likely to be

¹⁰ In the five separate robustness models run (Appendix H), the poverty variable's coefficient is of comparable size, the associated standard errors with the measure are very similar, and it is significantly associated with the decision to bribe in all but one regression (government effectiveness). This suggests that poverty played a similar role in most regressions and that its inclusion in the models did not unduly influence the estimated influence of exposure to the two treatment groups (local and religious) for which there was an imbalance.

asked to pay a bribe on a regular basis (Peiffer and Rose 2016). The significance of poverty in our regressions may therefore reflect the effect of socialization into a norm of bribe payment as much as socioeconomic status (Smith 2008).

Because the findings in Table 1 are indicative of the expectations articulated in the literature around corruption as a collective-action problem (H1) but do not fully confirm it, we test this hypothesis next.

Do Messages Influence Pessimistic Perceivers Differently?

Drawing on insights from the literature on social norms and corruption as a collective-action problem, H1 expects that all anticorruption messaging, regardless of framing, encourages bribery among those who already think that corruption is widely practiced in society. This is a conditional hypothesis. It suggests any messaging at all may encourage bribery, conditional on a preexisting high level of perceived corruption. Statistically speaking, conditional hypotheses like this are best tested with the introduction of an interaction term, which articulates when certain conditions between two variables are met. In this case, we created five interaction terms, which articulate exposure to each message and whether a person is what we call a pessimistic perceiver or had a preexisting perception that corruption is very widely practiced.

Identifying Pessimistic Perceivers

To identify pessimistic perceivers, we constructed an index based on the responses to three questions in our survey; how widespread corruption is in Nigeria, how common corruption is among public officials, and whether most ordinary people known to participants pay bribes. The response options to these questions were based on either four- or five-point scales.¹¹ We used principal component factor analysis to construct our index, extracting the common variance from responses to these three questions to calculate a single pessimism score for each individual. The factor analysis formed a single factor—our pessimism index—with an eigenvalue over the 1.0 threshold (1.39).

The index, which reflects generalized perceptions of how widespread corruption is practiced, ranges from -4.11 to 1.11, where a low score reflects disagreement that corruption is common or widespread and a high score reflects strong agreement that it is. The mean score for our sample is 0.00 with a majority—57% of the sample—scoring higher than the mean. To give a sense of what a score above a zero means, as it relates to the index's constituent terms, 99% of respondents that scored higher than a zero on our index agreed that corruption is common among public officials (19%, “common”; 80%, “very common”). Ninety-six percent of these respondents also agreed that most people they know have paid a bribe (43%, “agree”; 53%, “strongly

¹¹ For the wording and response options, see Appendix E.

agree”), and 98% of them rate corruption as being at least very widespread in Nigeria (11%, “very widespread”; 88%, “extremely widespread”).

Significantly, separate regression analyses (found in Appendix F) demonstrate that these perceptions of corruption were not influenced by exposure to the messages. This allows us to treat these perceptions as reflecting preexisting beliefs about the pervasiveness of corruption and our measure of pessimism as exogenous to the experimental conditions in our analyses. This finding is not surprising; others have similarly found that exposure to corruption messaging does not influence ideas about how widely practiced corruption is in society (Chong et al. 2015; Peiffer 2018).

Testing the Interactions

To test what effect the interaction between perceptions of widespread corruption and exposure to messages have on bribery in our game, we estimated a new logistic regression. For this we used the regression model reported in Table 1 as a base model and added to it the five interaction variables mentioned (exposure to each treatment \times perceptions of widespread corruption). Estimating the effect that an interaction variable has on a binary decision—like the choice to pay a bribe in our game—within a logistic regression is not as straightforward as it is in a linear regression when the dependent variable is continuous. Statistical inferences cannot be made about the association of an interaction term and a dependent variable by the coefficient and statistical significance registered from a standard logistic regression output (Ai and Norton 2003). This is because, for example, an insignificant and positive coefficient reported in a regression output may hide the fact that the interaction term is significantly associated with the dependent variable at low levels of both constituent terms but not at middle or higher levels.

Therefore, it is essential to scrutinize the effects of an interaction. For ease of interpretation, we focus on two categories of respondents—those who scored highly on our index of perceived corruption (the pessimistic perceivers) and those who did not (the nonpessimistic perceivers). To create these categories, we split the sample based on our pessimistic perceiver index. About the same percentage of people answered that corruption is extremely widespread in Nigeria (62%) and that corruption was very common among public officials (64%). So, using the zero position on the index, we divided the sample into a similar distribution between those who scored below zero on the factor index (43%, $n = 1,092$)—nonpessimistic perceivers—and those who scored higher than zero (57%, $n = 1,480$)—pessimistic perceivers.¹² It is tempting to think that individuals in the latter category are not actually “pessimistic” but simply “realistic” given the extent of corruption described above, but the survey we conducted

alongside the bribery game suggests that these individuals likely think the situation is worse than reality. While 96% of pessimists agree that most people they know have paid a bribe, agreement drops to 59% for nonpessimists. Against this, surveys have consistently found that a minority of Nigerians pay a bribe over a 12-month period (e.g., Pring and Vrushi 2019).

We report the estimated influence of messaging on the predicted probability of paying a bribe in our game for each group. Therefore, our results (Figure 2) reveal whether and how the effect of anticorruption messaging is moderated by prior beliefs about corruption.¹³ Almost across the board we find consistent results that strongly support H1, and this finding holds for different specifications of pessimistic and nonpessimistic perceivers as well as for examining the influence of exposure to each message in separate logistic regressions (see Appendix H for robustness checks). For pessimistic perceivers, exposure to four of the five messages significantly encouraged bribery in our game. Exposure to the fifth message (*local fight*), is positively associated with paying a bribe in the game among pessimistic perceivers, though this association is not significant at $p < 0.10$.

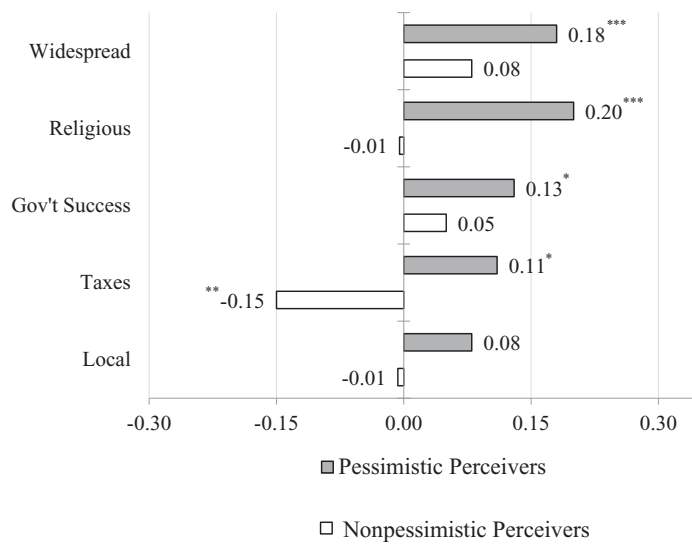
The messages also had a particularly strong substantive effect for this group. More specifically, among pessimistic perceivers, exposure to the *widespread* and *religious* messages increased the predicted probability of paying a bribe in our game by about 20 percentage points, respectively, while exposure to the *government success* and *taxes* messages increased the predicted probability of paying a bribe by an average of 12 percentage points. Instead of discouraging bribery, almost every message tested did the opposite for pessimistic perceivers.

These results support the hypothesis that exposure to any message about corruption may increase willingness to engage in corrupt practices among pessimistic perceivers. In turn, these findings directly support the growing body of research that shows that by inadvertently priming an issue a message may prompt unintended behavioral reactions. Our results also have important implications for policy interventions, showing that awareness-raising efforts may *encourage* corruption among those who already believe that corruption is a widespread problem. The real-world effect of these processes is significant because around two-thirds of Nigerians believe that corruption is widespread, including among public officials. Our findings therefore suggest that traditional anticorruption campaigns may increase the willingness of most of the population to pay bribes.

For nonpessimistic perceivers the story is starkly different; none of the messages are found to have encouraged bribery in our game. In other words, nonpessimistic perceivers who were shown the *widespread*, *religious*, *government success*, and *local fight* messages were just as likely to bribe as those who were not shown

¹² In the sample of respondents who played the bribery game (1,200), the distribution is the same: 43% ($n = 515$) as nonpessimistic perceivers and 57% ($n = 685$) as pessimistic perceivers.

¹³ We report the full regression models associated with Figure 2 in Appendix G.

FIGURE 2. Effect of Messaging on Bribery Payment among Pessimistic Perceivers and Nonpessimistic Perceivers

Note: Shift in predicted probability of bribing associated with message exposure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

a message at all (*control* group). This is not especially encouraging from a policy perspective either; it suggests that messages like these may not be influential, even among this “best case scenario” group. However, these null results do demonstrate that, at the very least, anticorruption messaging may not backfire for everyone.

Moreover, there is one important exception to this bleak story that may point the way to improving the effect of messaging in the future—namely, the effect that the *taxes* message has on bribery in our game among nonpessimistic perceivers. In this case—and this case only—we find that exposure to a message may discourage bribery. Indeed, for nonpessimistic perceivers exposure to the *taxes* message is associated with a sizeable 15-percentage-point decrease in the probability of bribing. This finding is significant for a number of reasons. First, it supports the idea that messages that connect corruption to the personal financial position of individual citizens can be more effective at reducing bribery, if only for this subgroup. Second, it provides indirect support for the broader argument that taxes and tax payment play a critical role in the evolution of a social contract in which citizens come to be more demanding of government and less tolerant of graft (Cheeseman and de Gramont 2017). Taken together, these implications suggest that building a wider direct tax base and emphasizing the connection between corruption and tax payment can help to build public support for a clean government.

This finding does not, however, lead to easy policy prescriptions, because the *taxes* message is only effective in our study when targeted at the little more than two-fifths of citizens who are not pessimistic about corruption. In turn, this raises the thorny question of

how to target this specific group in a way that would not simultaneously increase the willingness of most citizens to pay bribes—a point to which we return in the conclusion.

What is Associated with Pessimism?

Finding that the effect of anticorruption messages is conditional on a respondent’s perceptions of corruption raises the question of who is most likely to be a pessimistic perceiver and therefore react negatively. To examine this, an additional logistic regression was run in which being a pessimistic perceiver is treated as the dependent variable and the following variables as potential independent variables: gender, age, education, poverty level, level of interest in politics, whether the respondent is a supporter of the ruling party, whether they are a public employee, and whether they have personally witnessed an act of corruption in the last year (see Appendix I for details). To be clear, this examination is able to establish whether certain traits are associated with pessimism but not necessarily what *causes* pessimism.

Table 2 reveals that only two of the variables tested are associated with pessimism. One is gender, but it is important to note that the association between gender and pessimism is not particularly strong ($p = 0.08$) and has a limited substantive effect. After controlling for the effects of other variables, men are estimated to have a 5-percentage-point greater likelihood of being a pessimistic perceiver than are women. Recent studies focusing on gendered vulnerability to bribery suggest that, in general, men are more likely to bribe than women (Kukutschka and Vrushi 2019; Vrushi 2020),

TABLE 2. Associations with Being a Pessimist Perceiver

	<i>b</i>	SE	<i>p</i>
Female	−0.21	0.123	0.082
Education	−0.09	0.097	0.335
Poverty	−0.06	0.071	0.367
Age	−0.00	0.006	0.846
Political interest	−0.02	0.058	0.711
APC	0.12	0.157	0.451
Public employee	−0.03	0.284	0.912
Witness corruption	0.65	0.132	0.000
Constant	0.41	0.439	0.350
<i>N</i>	1,175		
Pseudo <i>R</i> ²	0.02		
Likelihood ratio	32.91		
Prob > χ^2	0.000		

Note: Displayed across the columns are coefficients (*b*), standard errors (SE), and *p* values.

and so it may be their greater vulnerability to bribery patterns that explains their greater pessimism.

The second significant association, which has a stronger relationship with pessimism ($p = 0.000$) and is more substantively important, is whether a respondent has personally witnessed corruption in the last year. While controlling for the effects of the other variables, those who witnessed corruption are estimated to have an 18-percentage-point greater likelihood of being a pessimistic perceiver than those who have not. Importantly, the combination of this finding with the results presented in Figure 2 suggests that individuals within this pessimistic-perceiving group may be locked in a vicious cycle: Those who witness corruption personally are more likely to hold pessimistic views of how pervasive it is in society. When confronted with anticorruption messaging, this pessimism may fuel resignation to corrupt environments and thus encourage corruption rather than the indignation such efforts are designed to provoke.

CONCLUSION

This paper has advanced the literature on the merits and dangers of anticorruption messaging in a number of ways. Our participants played a game in which they stood to make real money by paying a bribe, a methodology that offers a stronger—and more realistic—test of how individuals are likely to behave in the real world. We also tested the effect of a wider range of messages, including the effect of stressing the direct effect of corruption on individuals in terms of the taxes they pay. Moreover, we analyzed whether anticorruption messages have a differential effect on citizens.

In line with existing literature, the anticorruption messages we tested either had no effect or actually made Lagosians more likely to pay a bribe in our game.

We also demonstrate that the problematic consequences of anticorruption messages are not universal. More specifically, we have highlighted a particularly problematic vicious cycle that takes hold among what we have termed pessimistic perceivers. This group has come to see corruption as extremely widespread—which our findings suggest may be because they had a higher likelihood of witnessing it—and are likely to respond to anticorruption messages by becoming more willing to bribe. The identification of this vicious cycle and the specific individuals that it is most likely to affect demonstrates an important mechanism through which anticorruption collective action problems become self-reinforcing. One of the “cures” that has been most commonly promoted as a means of reducing corruption may not only have no positive effect but actually may compound the problem. Although it is important to be careful in generalizing beyond the case of Lagos, research on citizen perceptions has come to similar conclusions in Costa Rica, Papua New Guinea, and South Africa. Therefore, it seems likely that our findings reflect a broader pattern, especially in diverse urban areas that, like Lagos, feature significant levels of corruption—which is the majority of capital cities in the developing world (Williams and Dupuy 2018).

More broadly, our research highlights the way that priming individuals to think about problematic social processes can exacerbate existing collective action problems, and so it highlights the risk of negative unintended consequences for policy interventions in this area. Unfortunately, this challenge extends well beyond the field of corruption. Early research on government and donor messaging designed to reduce gender inequality reveals that it can foster problematic gender norms (Keller, Wilkinson, and Otjen 2010). This presents policy makers with something of a conundrum: the more they talk about major global challenges, the harder some of them are to resolve. There is no easy way out of this bind, but urgent action is required. At present, international donors, civil society groups, and governments seeking to affect attitudinal change appear to be overly confident that their interventions are doing good or at the very least doing no harm, but a growing body of research demonstrates that this is not the case.

Responding to this challenge effectively will require policy makers to change tack and place the potential cost of priming individuals to think about social ills at the heart of their responses. As Rothstein (2011) has argued, one way to do this with anticorruption messages is to design more subtle strategies that can “nudge” individuals into action by using indirect messages and prompts that do not lead recipients to focus on the scale of the problem at hand. Even when this is done, however, it will still be advisable to target messages in a way that minimizes the risk that they reinforce the worst fears of pessimistic perceivers. In turn, this will require avoiding blanket advertising strategies such as newspapers and billboards and instead carefully directing messages at particular individuals and groups. In some countries, the availability of social media and in-depth data on personal attitudes and behaviors may

allow for messages to be individually targeted with great precision; in others, the best that can be done at present may be to avoid forms of media that are particularly likely to be consumed by the kinds of citizens that are most likely to be pessimists in a given context.

None of these changes will be easy. Identifying which messages have the greatest influence for the smallest risk will require further research, and operationalizing different types of campaign strategies will require additional resources and a fundamental shift in approach. Given the costs and effort involved in undertaking these reforms, they are likely to meet with considerable resistance, but they cannot be put off: without them anticorruption messaging will do more harm than good.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S0003055421001398>.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available at the American Political Science Review Dataverse: <https://doi.org/10.7910/DVN/JV8GBI>.

ACKNOWLEDGMENTS

We are grateful to the team at Practical Sampling International for working with us to conduct the fieldwork for this research. We also thank Pallavi Roy, Mushtaq Khan, Elizabeth David-Barrett, Bill Mishler, Amma Panin, and Jennifer Merolla for reviewing previous drafts and providing helpful feedback. The views presented in this article are those of the authors and do not necessarily reflect the UK government's official policies or the views of SOAS-ACE or other partner organizations.

FUNDING STATEMENT

This publication is an output of the SOAS Anticorruption Evidence (ACE) research consortium funded by UK aid from the UK Government as per contract [P0 7073].

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors declare the human subjects research in this article was reviewed and approved by the University of Birmingham's Ethical Review Committee (ERN_18-1233). The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research.

REFERENCES

- Ai, Chunrong, and Edward Norton. 2003 "Interaction Terms in Logit and Probit Models." *Economic Letters* 80 (1): 123–29.
- Afrobarometer. 2020. "Sampling Principles and Weighting." Cape Town: Afrobarometer. <http://afrobarometer.org/surveys-and-methods/sampling-principles>.
- Anduiza, Eva, Aina Gallego, and Jordi Munoz. 2013. "Turning a Blind Eye: Experimental Evidence of Partisan Bias in Attitudes towards Corruption." *Comparative Political Studies* 46 (12): 1664–92.
- Baez Camargo, Claudia. 2017. "Can a Behavioural Approach Fight Corruption?" *Basel Institute on Governance Policy Brief* No. 1. https://edoc.unibas.ch/66324/1/20181029115507_5bd6e70b67471.pdf.
- BBC. 2019. "Minimum Wage: Nigeria Don Finally Approve N30,000 as Workers Pay." BBC. March 19. <https://www.bbc.com/pidgin/47627269>.
- Berinsky, Adam, Vincent Hutchings, Tali Mendelberg, Lee Shaker, and Nicholas Valentino. 2011. "Sex and Race: Are Black Candidates More Likely to Be Disadvantaged by Sex Scandals?" *Political Behavior* 33 (2): 179–202.
- Bicchieri, Cristina, and Eugen Dimant. 2019. "Nudging with Care: The Risks and Benefits of Social Information." *Public Choice*, 1–22. <https://doi.org/10.1007/s11127-019-00684-6>.
- Bodea, Cristina, and Adrienne LeBas. 2016. "The Origins of Voluntary Compliance: Attitudes toward Taxation in Urban Nigeria." *British Journal of Political Science* 46 (1): 215–38.
- Brody, Richard, and Benjamin Page. 1972. "Comment: Assessment of Policy Voting." *American Political Science Review* 66 (2): 450–8.
- Carter, Travis, Melissa Ferguson, and Ran Hassin. 2011. "A Single Exposure to the American Flag Shifts Support toward Republicans up to 8 Months Later." *Psychological Science* 22 (8): 1011–8.
- Cheeseman, Nic. 2015. *Democracy in Africa: Successes, Failures, and the Struggle for Political Reform*. Cambridge: Cambridge University Press.
- Cheeseman, Nic, and Brian Klaas. 2018. *How to Rig an Election*. New Haven, CT: Yale University Press.
- Cheeseman, Nic, and Caryn Peiffer. 2021. "Replication Data for: The Curse of Good Intentions: Why Anticorruption Messaging Can Encourage Bribery." Harvard Dataverse. Dataset. <https://doi.org/10.7910/DVN/JV8GBI>.
- Cheeseman, Nic, and Diane de Gramont. 2017. "Managing a Mega-City: Learning the Lessons from Lagos." *Oxford Review of Economic Policy* 33 (3): 457–77.
- Cheeseman, Nic, Gabrielle Lynch, and Justin Willis. 2020. *The Moral Economy of Elections in Africa: Democracy, Voting, Virtue*. Cambridge: Cambridge University Press.
- Chong, Alberto, Ana De La O, Dean Karlan, and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification." *The Journal of Politics* 77 (1): 55–71.
- Corbacho, Ana, Daniel Gingerich, Virginia Oliveros, and Mauricio Ruiz-Vega. 2016. "Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica." *American Journal of Political Science* 60 (4): 1077–92.
- Dewan, Torun, Macartan Humphreys, and Daniel Rubeson. 2014. "The Elements of Political Persuasion: Content, Charisma, and Cue." *The Economic Journal* 124 (574): 257–92.

- Druckman, James. 2001. "The Implications of Framing Effects for Citizen Competence." *Political Behaviour* 23 (3): 225–56.
- Economist. 2011. "Nigeria's Business Capital: A Rare Good Man." *Economist*, May 5. <https://www.economist.com/middle-east-and-africa/2011/05/05/a-rare-good-man>.
- Erisen, Cengiz, Milton Lodge, and Charles Taber. 2014. "Affective Contagion in Effortful Political Thinking." *Political Psychology* 35 (2): 187–206.
- Hoffman, Leena Koni, and Raj Navanit Patel. 2017. "Collective Action on Corruption in Nigeria: A Social Norms Approach to Connecting Society and Institutions." Report, Chatham House, May 17. <https://reader.chathamhouse.org/collective-action-corruption-nigeria-social-norms#>.
- Incerti, Trevor. 2020. "Corruption Information and Vote Share: A Meta-Analysis and Lessons for Experimental Design." *American Political Science Review* 114 (3): 761–74.
- Jacoby, William. 2000. "Issue Framing and Public Opinion on Government Spending." *American Journal of Political Science* 44 (4): 750–67.
- Johnson, Jesper, Nils Taxell, and Dominik Zaum. 2012. "Mapping Evidence Gaps in Anticorruption: Assessing the State of the Operationally Relevant Evidence on Donors' Actions and Approaches to Reducing Corruption." U4 Issue 2012 (7). <http://www.cmi.no/publications/publication/?4624=mappingevidence-gaps-in-anticorruption>.
- Jones, Alun. 2011. "Awareness-Raising Campaigns." Vienna: United Nations Office on Drugs and Crime, August 22–24. https://www.unodc.org/documents/treaties/UNCAC/WorkingGroups/workinggroup4/2011-August-22-24/Presentations/Alun_Keith_Jones_UNODC_-_Communications_and_Advocacy.pdf.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Keller, Sarah, Timothy Wilkinson, and A. J. Otjen. 2010. "Unintended Effects of a Domestic Violence Campaign." *Journal of Advertising* 39 (4): 53–68.
- Khan, Mushtaq, Antonio Andreoni, and Pallavi Roy. 2019. "Anticorruption in Adverse Contexts: Strategies for Improving Implementation." SOAS-ACE Working Paper 13. <https://eprints.soas.ac.uk/23495/>.
- Kobis, Nils, Marleen Troost, Cyril Brandt, and Ivan Soraperra. 2019. "Social Norms of Corruption in the Field: Social Nudges on Posters Can Help to Reduce Bribery." *Behavioural Public Policy* December: 1–28. doi:10.1017/bpp.2019.37.
- Krosnick, Jon, and Donald Kinder. 1990. "Altering the Foundations of Support for the President through Priming." *American Political Science Review* 84 (2): 497–512.
- Kukutschka, Roberto, and Jon Vrushy. 2019. *Global Corruption Barometer: Middle East & North Africa 2019*. Berlin: Transparency International.
- Lenz, Gabriel. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53 (4): 821–37.
- Lupia, Arthur. 1994. "Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review* 88 (1): 63–76.
- Manning, Mark. 2009. "The Effects of Subjective Norms on Behaviour in the Theory of Planned Behaviour: A Meta Analysis." *British Journal of Social Psychology* 48 (4): 649–705.
- Marquette, Heather. 2012. "'Finding God' or 'Moral Disengagement' in the Fight against Corruption in Developing Countries? Evidence from India and Nigeria." *Public Administration and Development* 32 (1): 11–26.
- Mattes, Robert, Michael Bratton, Yul Davids, and Cherrel Africa. 2000. "Public Opinion and the Consolidation of Democracy in Southern Africa." Afrobarometer Working Paper 7. <http://afrobarometer.org/sites/default/files/publications/Working%20paper/AfropaperNo7.pdf>.
- Moore, Mick, Wilson Prichard, and Odd-Helge Fjeldstad. 2018. *Taxing Africa: Coercion, Reform and Development*. London: Zed Books.
- Mungiu-Pippidi, Alina. 2011. "Contextual Choices in Fighting Corruption: Lessons Learned." NORAD Report 4/2011, July. Oslo: Norwegian Agency for Development Cooperation. <https://norad.no/en/toolspublications/publications/2011/contextual-choices-in-fighting-corruption-lessons-learned/>.
- Nosek, Brian, Jesse Graham, and Charlie Hawkins. 2010. "Implicit Political Cognition." In *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*, eds. Bertram Gawronski and B. Keith Payne, 548–64. New York: The Guilford Press.
- Olivier de Sardan, Jean-Pierre. 1999. "A Moral Economy of Corruption in Africa?" *Journal of Modern African Studies* 37 (1): 25–52.
- Peiffer, Caryn. 2017. "Getting the Message: Examining the Intended – and Unintended – Impacts of Corruption Awareness-Raising." Developmental Leadership Program, August 2017. <https://www.dlprog.org/publications/research-papers/getting-the-message-examining-the-intended-and-unintended-impacts-of-corruption-awareness-raising>.
- Peiffer, Caryn. 2018. "Message Received? Experimental Findings on How Messages about Corruption Shape Perceptions." *British Journal of Political Science* 50 (3): 1207–15.
- Peiffer, Caryn, and Grant Walton. 2019. "Overcoming Collective Action Problems through Anticorruption Messages." *Development Policy Centre Discussion Paper No. 77*. http://devpolicy.org/publications/discussion_papers/DP77-Overcoming_collective_action_problems.pdf.
- Peiffer, Caryn, and Linda Alvarez. 2016. "Who Will Be the Principled Principals? Examining Willingness to Actively Oppose Corruption." *Governance* 29 (3): 351–69.
- Peiffer, Caryn, and Richard Rose. 2018. "Why Are the Poor More Vulnerable to Bribery in Africa? The Institutional Effects of Services." *Journal of Development Studies* 54 (1): 18–29.
- Persson, Anna, Bo Rothstein, and Jan Teorell. 2013. "Why Anticorruption Reforms Fail: Systemic Corruption as a Collective Action Problem." *Governance* 26 (3): 449–71.
- Persson, Anna, Bo Rothstein, and Jan Teorell. 2019. "Getting the Basic Nature of Systemic Corruption Right: A Reply to Marquette and Peiffer." *Governance* 32 (4): 799–810.
- Prichard, Wilson. 2015. *Taxation, Responsiveness and Accountability in Sub-Saharan Africa: The Dynamics of Tax Bargaining*. Cambridge: Cambridge University Press.
- Pring, Coralie, and Jon Vrushy. 2019. "Global Corruption Barometer Africa 2019. Citizens Views and Experiences of Corruption." Transparency International, July 11. https://images.transparencycdn.org/images/2019_GCB_Africa3.pdf.
- Reinikka, Ritva, and Jakob Svenson. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 95 (7–8): 956–66.
- Richards, Kathy. 2006. *What Works and Why in Community-Based Anticorruption Programs*. Melbourne: Transparency International Australia.
- Riker, William. 1986. *The Art of Political Manipulation*. New Haven, CT: Yale University Press.
- Rocha Menocal, Alina, and Nils Taxell. 2015. "Why Corruption Matters: Understanding Causes, Effects, and How to Address Them." Evidence Paper. London: UK Department for International Development. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/406346/corruption-evidence-paper-why-corruption-matters.pdf.
- Rose-Ackerman, Susan. 2015. "Are Corrupt Elites Necessary for Corrupt Countries?" In *Elites, Institutions and the Quality of Government: Executive Politics and Governance*, eds. Carl Dahlström and Lena Wängnerud, 33–47. London: Palgrave Macmillan.
- Rothstein, Bo. 2011. "Anticorruption: The Indirect 'Big Bang' Approach." *Review of International Political Economy* 18 (2): 228–50.
- Rothstein, Bo. 2013. "Corruption and Social Trust." *Social Research* 80 (4): 1009–32.
- Rothstein, Bo, and Daniel Eek. 2009. "Political Corruption and Social Trust: An Experimental Approach." *Rationality and Society* 21 (1): 81–112.
- Smith, Daniel. 2008. *A Culture of Corruption: Everyday Deception and Popular Discontent in Nigeria*. Princeton, NJ: Princeton University Press.

- Sniderman, Paul, and Sean Theriault. 2004. "The Structure of Political Argument and the Logic of Issue Framing." In *Studies in Public Opinion: Attitudes, Non-Attitudes, Measurement Error, and Change*, eds. William Saris and Paul Sniderman, 133–65. Princeton, NJ: Princeton University Press.
- Tankard, Margaret, and Elizabeth Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10 (1): 181–211.
- Transparency International. 2020. "Corruption Perception Index, 2019." Transparency International, Berlin. <https://www.transparency.org/cpi2019>.
- Uganda Media Centre. 2020. "ACU Chief Rallies Religious Leaders in the Fight against Corruption." September 9. <https://www.mediacentre.go.ug/media/acu-chief-rallies-religious-leaders-fight-against-corruption>.
- United Nations. 2004. "United Nations Convention against Corruption." Vienna: United Nations Office on Drugs and Crime. https://www.unodc.org/documents/treaties/UNCAC/Publications/Convention/08-50026_E.pdf.
- United Nations Office on Drugs and Crime (UNODC) Research. 2019. "Corruption in Nigeria: Patterns and Trends." UNODC. December 2019. https://www.unodc.org/documents/data-and-analysis/statistics/corruption/nigeria/Corruption_in_Nigeria_2019_standard_res_11MB.pdf.
- Van Duyn, Emily, and Jessica Collier. 2019. "Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media." *Mass Communication and Society* 22 (1): 29–48.
- Vrushi, Jon. 2020. "Global Corruption Barometer: Asia 2020." Berlin: Transparency International.
- Williams, Aled, and Kendra Dupuy. 2018. "Corruption and the City: How Aid Donors Can Support Integrity Building in Urban Spaces." *U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute U4 Brief* 2018: 2. <https://www.u4.no/publications/corruption-and-the-city>.
- Winters, Mathew, and Rebecca Weitz-Shapiro. 2013. "Lacking Information or Condoning Corruption: When Will Voters Support Corrupt Politicians?" *Comparative Politics* 45 (4): 418–36.