doi: https://doi.org/10.15446/rcp.v33n2.103188

Transparency and Altruistic Punishment in an Experimental Model of Cooperation to Corruption Through Economic Games

Transparency and punishment against corruption

JUAN JOSÉ DUARTE-BARROSO
University of Guanajuato, León, México

CHRISTIAN ENRIQUE CRUZ TORRES
University of Guanajuato, León, México

How to cite this article: Duarte-Barroso, J. J. & Cruz, C. E. (2024). Transparency and altruistic punishment in an experimental model of cooperation to corruption through economic games. *Revista Colombiana de Psicología, 33 (2)*, 43-62. https://doi.org/10.15446/rcp.v33n2.103188

This research received funding from the Institutional Call for Scientific Research 2021 (project 158/2021), as well as a scholarship from CONACYT (764271).

Correspondence: Christian Enrique Cruz Torres (https://orcid.org/0000-0002-4286-4697); Department of Psychology, University of Guanajuato León Campus. Address: Blvd. Puente Milenio No. 1001 Fracciones del Predio de San Carlos, León, Guanajuato, México, C.P. 37670. Telephone: +52 477 2674900, Ext. 3692. Email: christian.cruz@ugto.mx

SCIENTIFIC RESEARCH ARTICLE

RECEIVED: June 212t, 2022 - ACCEPTED: September 25th, 2023

Transparency and altruistic punishment in an experimental model of cooperation to corruption through economic games

Abstract

This work integrates cooperation, punishment, treasury damage, and norms transgression in three variants of a single experimental model of corruption. Participants formed words with predetermined letters, receiving a reward for each word, besides an individual reward taken from the common fund if they reached the goal. A manipulation in the letters made it impossible to reach the goal, so reporting exceeding it implied cheating for a benefit. Three studies model the effects of signaling, descriptive norms, and the possibility of punishing or investigating corruption acts (transparency). 248 participants were randomly assigned to the conditions of each study. Significatively less cheating behavior was found in reports of words and earnings in Studies 1 and 3, but not in Study 2. The experimental model reveals the potential of transparency as an alternative to diminishing corruption with less social cost than altruistic punishment. The relevance of these results for implementing public policies was discussed.

***Keywords*:** Altruistic punishment, corruption, descriptive norms, signaling, transparency.

Transparencia y Castigo Altruista en un Modelo Experimental de Cooperación a la Corrupción a Través de Juegos Económicos

Resumen

Este trabajo integra la cooperación, el castigo, el daño al erario y la transgresión de normas en tres variantes de un único modelo experimental de corrupción. Los participantes formaban palabras con letras predeterminadas, recibiendo una recompensa por cada palabra, además de una recompensa individual tomada del fondo común si alcanzaban la meta. Una manipulación en las letras hacía imposible alcanzar la meta, por lo que informar de que se superaba implicaba hacer trampas para obtener un beneficio. Tres estudios modelan los efectos de la señalización, las normas descriptivas y la posibilidad de castigar o investigar los actos de corrupción (transparencia). 248 participantes fueron asignados aleatoriamente a las condiciones de cada estudio. Se encontró un comportamiento de engaño significativamente menor en los informes de palabras y ganancias en los Estudios 1 y 3, pero no en el Estudio 2. El modelo experimental revela el potencial de la transparencia como alternativa para disminuir la corrupción con menor coste social que el castigo altruista. Se discute la relevancia de estos resultados para la implementación de políticas públicas.

Palabras clave: Castigo altruista, corrupción, normas descriptivas, señalización, transparencia.

According to Transparency International (2016), in 2015 Mexico obtained 35 out of 100 points in its corruption perception index, ranking as one of the most corrupt countries evaluated. This rating steadily worsened from that date until the 2020 measurement where it reached just 31 points, placing it below the average score for the American continent (43) and other regions as Sub-Saharan Africa (32) (Transparency International, 2021). These corruption levels affect the economy. For example, Dang et al. (2022) show that corruption increases the informal economy, and Spyromitros and Panagiotidis (2022) show that high levels of corruption, added to bureaucratic inefficiency, hinder the economic growth of countries.

Traditionally, corruption has been studied as a problem for public officials and their institutions. Treisman (2000), for example, defines corruption as the abuse of public office for personal gain. However, the levels of corruption observed in Mexico cannot be explained without the participation, or at least the consent, of a large part of the population, indicating a severe problem of respect for legality. In data from INEGI (2015), 22% of those surveyed state that they have paid bribes in procedures to establish a company, 23% for operations before the public prosecutor, and 55% for matters related to general security authorities. These high rates of citizen participation in the face of corruption may point to a normalization of corruption. For example, the Constitutional Culture Surveys in Mexico (Fierro et al., 2011) show that 41% of citizens would be willing to violate the law if they consider they are correct, while 21% declare that they agree or strongly agree with the phrase “breaking the law is not so bad, the bad thing is that they catch you” (Fierro et al., 2017).

From these antecedents, the definition of Treisman (2000), centered on public officials as sole agents of corruption, has been exceeded. Sutherland (1940) defines corruption as a violation of delegated or implicit trust from a broader perspective. From a more operational approach, Transparency International (2019) defines corruption as the abuse of entrusted power for private gain, classifying it as large (which occurs in the upper echelons of government), political (exercised by decision-making officials to modify procedures of the allocation of resources for their benefit), and less (exercised by public officials in their daily interactions with ordinary citizens). From this last category, corruption is a problem whose understanding and solution require including citizens as agents, not only as victims.

Persson et al. (2013) conceive corruption as a collective action problem, where people act according to the behavior they expect from others. Mungiu-Pippidi (2013) considers that collective action can foster an ethical universalism that allows for reaching a balance of social well-being. However, Marquette and Peiffer (2015) consider this approach to fighting corruption incomplete as it does not include the possibility of monitoring each other to cooperate for the collective good.

Cooperation, common goods, and corruption

Cooperation is a practice where an individual or group puts part of their resources (e.g., time, money, work) into a joint task with another individual or group to obtain a common benefit (Bowles & Gintis, 2011). In cooperation, a cooperator is identified as someone who pays a cost so that the parties involved obtain a benefit, and a free-rider is someone who does not pay that cost and shares the benefits anyway (Nowak, 2006).

Although evolution implies competition between individuals, where selfish behaviors are rewarded, cooperation is necessary to build new levels of organization (Nowak, 2006). These more complex levels of social organization give rise to collective goods, which differ in their capacity for exclusion due to their magnitude and nature. For example, you can deny the access easily to someone who refuse to pay a movie ticket, but its harder to exclude from living in a secure neighborhood to someone that refuse to pay its taxes as the rest of neighbors does. They also differ in their level of exploitation by use. For example, when a car is purchased, it is no longer available to someone else, while when public transportation is used, it is still available to others (Ostrom, 2003, 2010).

For this research, the interest is in the common-pool resources, that refer to natural or human-made resources and are considerably large enough to make the exclusion of their use or benefit feasible (Ostrom, 2011), and whose consumption or exploitation reduces more notably the total of goods available to others (Ostrom, 2003), but they have evident and quantifiable decline due to their exploitation; for example, timber forests, water, and the public fund. Because of their deterioration and the difficulty of excluding free-riders, common-pool resources are vulnerable to overuse, which can deplete resources and destroy the ability of the system to restore itself. This phenomenon, known as the tragedy of the commons (Hardin, 1968; Ostrom, 2011), results when people act only according to their immediate personal benefits and make excessive use of a scarce resource until it is exhausted, finally leading all to ruin (e.g., deforestation, bankruptcy).

Weisel and Shalvi (2015) investigated corruption through a game where two players roll dice in several rounds without supervision and receive a reward in each round if both obtain the same number. The results show that, even without the possibility of communicating or agreeing previously, the players report coincidences superior to those expected by chance, pointing out the existence of cooperation between individuals as a critical element of corruption, in this case, cheating to get more profit. This is a relevant finding, but the experimental procedure does not entirely model corruption, as the resources obtained by the players come from the experimenters and not from a common fund fed by contributions from the players, as in the case of a treasury provided by the collection of taxes.

In another study, Fehr and Gächter (2000) designed a procedure where players had to decide whether to contribute part of their profits from each round to a fund that would be shared equally among all at the end or keep all their income. During some rounds, the players only decided whether they contributed to the common fund. Still, in another series of rounds, they could pay for the experimenters to punish those who had not contributed to that round, an action known as altruistic punishment since it implies a cost but not a direct and immediate benefit. This possibility of punishing free riders steadily increased cooperation, showing the power of altruistic punishment, originated by the vigilance of the players (citizenship) and not of the experimenters (authority or government) to reduce desertion significantly. However, this experimental procedure does not entirely model corruption either, since the alternatives of cooperating or not in the common fund are validated by being included in the rules of the game, while norms in society dictate that the correct thing is to contribute to the common fund and deserting is immoral and illegal, implying a clear transgression of the norm.

Furthermore, unlike the high disposition to altruistic punishment reported in the studies by Fehr and Gächter (2000, 2002), other studies show that people are very reluctant to punish those who have not directly insulted them (Pedersen et al., 2018), and this disposition is almost null in conditions outside the laboratory (Pedersen et al., 2020). The latter leads to think about alternative forms of participation, more affordable than the direct exercise of punishment, for example, anonymous reporting and the requirement of the authority to investigate and make transparent presumed acts of corruption (Bauhr et al., 2020).

Finally, the procedure of Fišar et al. (2016) includes three players, who occupy the roles of two citizens and an officer. Each participant received 100 coins at the beginning of the activities and was made aware that was playing with two other people and that each was given the same number of coins. The first citizen and the officer could keep their coins or cooperate and take 20 more coins each, which would be subtracted from the second citizen. In turn, this second citizen could just keep their 60 coins or pay ten coins to punish the officer and the other citizen, holding only 50 coins. This procedure incorporates altruistic punishment and cooperation as essential elements of corruption but has the weakness of validating the possibility of conspiring to be corrupt into the rules. Again, as in the Fehr and Gächter (2000) procedure, corruption behaviors were incorporated into the rules, validating them as an option, as occurs, for example, in baseball, where players can steal a base without violating the rules of the game, even if it imply a higher risk.

As in the procedure by Fišar et al. (2016), acts of corruption typically require the cooperation of two or more people, for example, a citizen and an official, in the case of bribery and other forms of petty corruption. As these acts are punishable by law for both citizens and officials, the risks of cooperation are compounded by the risk that the counterpart denounces, which in some laws implies the benefits of reduction of penalties for the whistleblower (Piccolo & Immordino, 2017). Cooperation requires a coordination process, where the parties involved must successfully communicate their intentions to coincide with their shared interests (Bacharach, 2018). This is complicated for acts of corruption, considering that proposing someone else to participate in the act of corruption implies the risk of being denounced or exposed, which, in the case of Mexico at least, has given rise to many covert forms of language (Legorreta & de Mola, 2016) that facilitate coordination and cooperation in these acts with less risk for the parties involved. For these reasons, the experimental procedure proposed here includes a manipulation where the participants receive signals from a confederate, reporting more words than she has formed, covertly inviting to falsify the results to obtain more profits from the common fund, which would model cooperation in the act of corruption.

Considering the analyzed antecedents, this work integrates into a new experimental procedure the strengths of the studies that experimentally model cooperation in unsupervised dishonest acts (Weisel & Shalvi, 2015), the covert invitation to cooperate in a corruption act (Fišar et al., 2016) and the use of public goods games and altruistic punishment to reduce free-riding (Fehr & Gächter, 2000, 2002), although in this case, to better model the existence of a treasury, the initial common fund was formed by resources contributed by the players.

First, we tested the experimental procedure in two studies by manipulating two variables whose effects on cooperation and corruption are known from previous studies. The first study analyzed signaling effects on cooperation for personal over collective benefit. The second study examines the impact of descriptive social norms on cooperation for personal gain. Finally, considering the low disposition to altruistic punishment (Pedersen et al., 2018, 2020), a third study compared altruistic punishment with another condition of request for transparency, where the participants can request the investigation and transparency of possible acts of corruption from the other players.

Then, this paper aims to explain the effects of signaling, social norms, and the possibility of punishment for cooperating in corruption. The hypothesis is that cooperation with corruption will be greater when signaling and the social norm of dishonesty are present, but there will be less cooperation with corruption when there is the possibility of sanction.

Study 1: signaling effects on cooperation for personal benefit

Signaling is when two parties have access to different information, so some have more information than others (Spence, 2002). Since this asymmetric information is private, people with the information could make better decisions than people who do not have it (Connelly et al., 2011). Then a person, *the sender*, has private information (*signal*) that can be positive or negative and can be offered to *the receiver*. The latter receives and interprets the information and sends a response to the sender. For the signaling to be successful, the sender must obtain benefits for some action that the receiver performs and that he would not have achieved without the information received (Banks & Sobel, 1987; Connelly et al., 2011). In a corruption act as bribery an official can offer a signal to a citizen, seeking to accept it, and both obtain a mutual benefit that would not be received if someone denounces (not cooperate); it is up to the citizen to get that signal and to cooperate with the official for mutual benefit.

Signaling has already been shown to have positive effects on cooperation. In a two-person strategy game, Salahshour (2019) found that, despite their apparent cost, signals evolve due to their capacity to elicit cooperation. Heinz and Schumacher (2017) examined the signs that the participants' curriculum reported on the willingness to cooperate as a team, finding that contributions in a public goods game increased following the degree of social participation indicated in the curriculum.

This study evaluates the effects of signaling on cooperation for personal gain in a situation that models petty corruption. The hypothesis is that higher cheating behavior will be found in the signaling condition compared to the control condition.

Method

Participants

The participants were 56 people from the city of León, Guanajuato, a predominantly urban region that is the seventh most populated in Mexico with 1.7 million inhabitants, whose economy is based mainly on the automotive and footwear industries. 57.14% are women, and 42.86% are men, with an average age of 20.79 (*SD* = 5.14). 7.15% had incomplete high school or lower studies, 39.29% completed high school, 44.64% had incomplete university studies, and 8.93% completed university studies. Participants were randomly assigned, 28 in control and 28 in the experimental condition. Of the 100 people invited, 19 people declined the invitation, 19 agreed to participate but did not continue with the communication, and only six people were excluded from the analysis for guessing at the study objectives.

Procedure

The inclusion criteria were being 18 years of age or older and being familiar with the use of instant messaging as Google Chat and Facebook Messenger, and as exclusion criteria having partial or completed university studies in psychology or economics, to avoid anticipating experimental procedures and manipulations by having a background in this type of research. Based on previous studies of corruption that show significant differences between the sexes (Rivas, 2013), it was sought to preserve the same proportion of men and women as much as possible.

Given the sanitary restrictions due to the COVID-19 pandemic, all contact with the participants and the experimental procedures were carried out by virtual means to avoid contagion risks to participants and experimenters.

Participants were invited through the research team's social networks, explaining that their participation would be voluntary, anonymous, and the data provided would be confidential, analyzed for scientific research purposes, and would be protected by the titular researcher, also informing them about the study objective, its procedures, and estimated duration. They were also told that to participate they had to contribute 5 Mexican pesos (approximately a quarter of a dollar) to form a common fund. Those who satisfied the inclusion and exclusion criteria and agreed to participate were sent a Google Forms link by email where they could read the informed consent in detail and, if they decided, express their consent by checking a box and continuing to respond to a sociodemographic data form. After answering the format, the description of the experimental procedure was presented as an online game where they had to form words together with other participants. They were reminded that they should have $5 on hand that they would contribute to creating the common fund. The informed consent also stated that in the game, they could win, but there was also the possibility of losing their initial contribution depending on their performance in the game. This helps to bring realism and relevance to their decisions, being important that participants felt they could win or lose. The experiment was presented to them as a study to measure the ability to form words, so, moral decisions were not the focus of the procedure. Furthermore, as an ethical consideration, the players received the game's winnings but did not pay their contributions or losses at the end of the procedure.

They received an email account and password to enter the Hangouts platform (https://hangouts.google.com), that prevented participants from knowing who they were playing with and controlling and maintaining a record of communication between them.

All participants were assigned the role of player two and the number of pair three, in each experimental session one participant and seven confederates played. Once on the Hangouts platform, a private chat with their partner and a group chat appeared with a link and instructions that directed them to the Socrative page (https://www.socrative.com). On this page, they were presented with the game instructions.

In the beginning, six supposedly random letters appeared on their screen. With these letters, both players in each pair had to form as many Spanish words as possible of at least four letters, for which they had one minute. They were asked to have a sheet at hand to write the words formed but only report the number of words, under the argument that it was easier to record what they won each round. The real reason was to generate a perception as it was easy to provide a false number of words. In each round, the letters appeared at the top of the screen, and the answer options indicated the number of words formed and the corresponding amount of money (e.g., one word = $.50 to the common fund, $0.125 per pair).

Both players earned $1 for each word they formed, which was accumulated in the common fund distributed among all at the end. For each round, a goal of 5 words was set. If the pair exceeded that goal, the players could conserve the profits of those rounds in a private fund and take $1 from the common fund. Except for the first two rounds, the letters provided were previously tested to limit the possible words to 5 (the letters used and the words formed with these can be seen in the Appendix A). This manipulation made it possible to know if the players reported more words than they could form.

In the experimental condition, in rounds six and nine, player one (confederate) proposed to player two (participant) to report more words to overcome the goal and obtain higher profits (signaling manipulation). In the control condition, the participants did not receive messages at all.

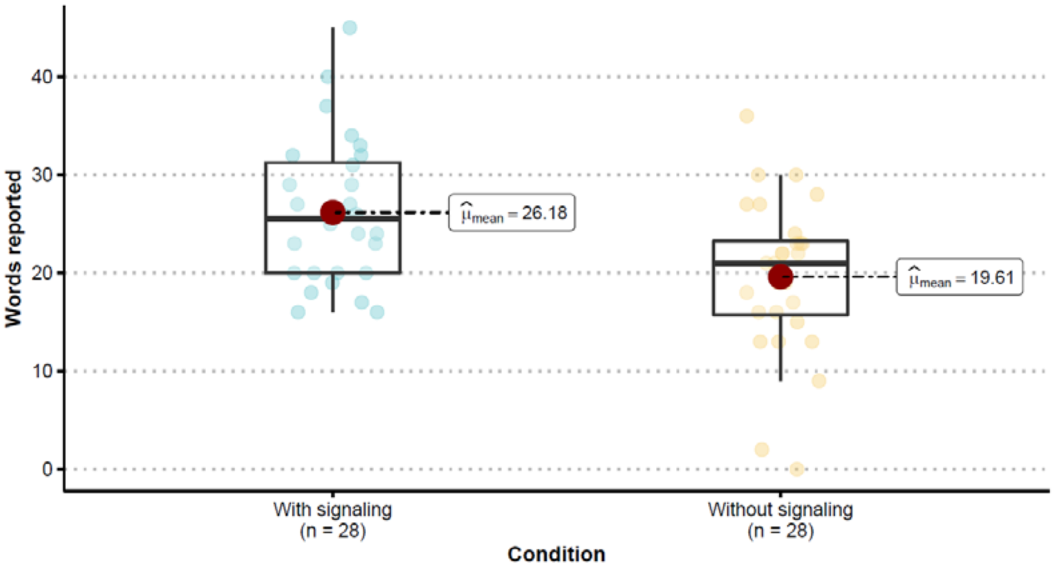
At the end of the ten rounds, the participants were informed that the session had concluded. The post-experimental interview was carried out to verify that they fully understood their decisions in the game and were not suspicious of the experimental manipulations. At that time, the common fund was distributed among all the participants, and was agreed on how to send them their corresponding earnings, indicating that they could also keep their initial contribution. Finally, they were informed about the experimental manipulations of the procedure, and about the real objectives, reiterating that their decisions would be confidential and would have no repercussions of any kind, and the contact details with the research team were reiterated for communication with doubts or any additional information. There were no complaints or comments of discomfort with the procedure. The procedure was reviewed and accepted by the Institutional Committee of Bioethics in Research of the Guanajuato University (CIBIUG-P16-2021).

Results

The analysis was carried out in the R software (R Core Team, 2021), the effect sizes were calculated with the *effectsize* package (Ben-Shachar et al., 2020), and the graphs in the *ggstatsplot* package (Patil, 2021).

Only 17.86% of the participants reported having exceeded the goal in the condition without signaling, while 50% did so in the condition with signaling, a difference statistically significant with medium effect size according to the chi-square test with Yates's adjustment $\chi^2(1) = 5.09$, $p = .023$, 1-β = .52, V = .31, 95% CI [0, 1]. In addition, using the Student's t-test it was found that the averages of the reported words were also higher in the condition with signaling (26.18, *SD* = 7.38) compared to the condition without signaling (19.61, *SD* = 8.02), statistically significant differences with large effect size, $t(54) = 3.19$, $p = .002$, 1-β = .50, $d = 0.85$, 95% CI [0.27, 1.41], as seen in Figure 1.

Figure 1
Words reported by participants according to the condition in which they played (Study 1)

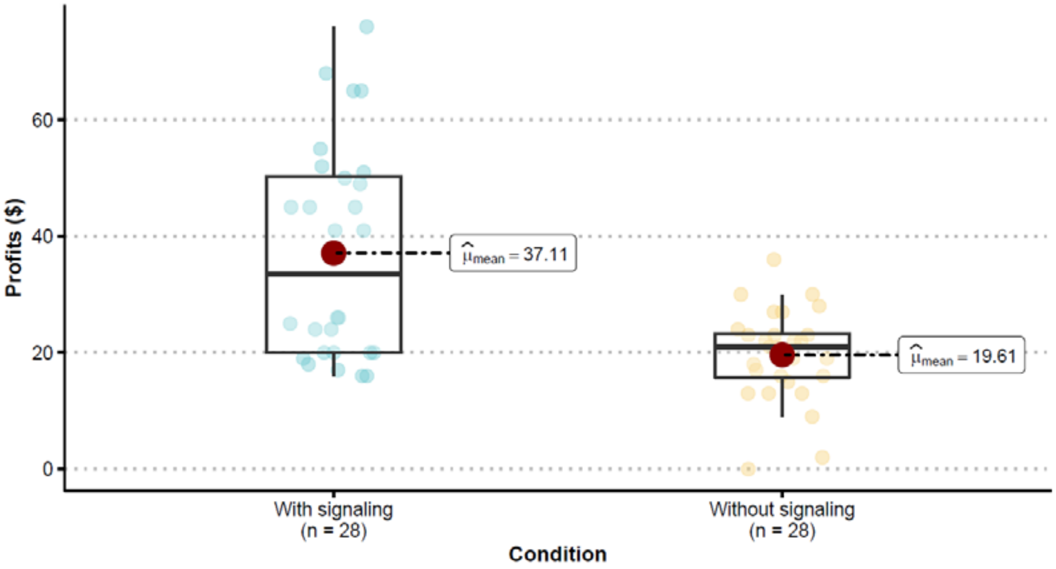


Note. Average of words reported, quartiles, and outliers in each condition

Additionally, the differences in profits in both conditions were analyzed. As indicated in the method, in rounds six and nine a confederate suggested that participants report more than six words and exceed the goal for higher profits. This is because, by reporting six or more words, the profits went to a private account that was not divided with the other pairs in the game, but only when both team members exceeded the goal. Therefore, the participants' reports of six words or more in rounds six and nine were multiplied by four, since those profits were not shared with the other three game pairs.

Through the Mann-Whitney-Wilcoxon test, significant differences with large effect size were found in the profits according to the condition in which they played, $W = 601$, $p < .001$, 1-β = .53, $r_b = 0.53$, 95% CI [.28, .72], being higher in the condition with signaling (*M* = 37.11, *SD* = 18.44), than in the condition without signaling (*M* = 19.61, *SD* = 8.01), as can be seen in Figure 2.

Figure 2
Average profits of participants according to the condition in which they played (Study 1)



Note. Average profits, quartiles, and outliers in each condition

Discussion

As expected in the hypothesis, the participants in the signaling condition reported exceeding the limit in more rounds than in the without-signaling condition; that is, they lied more to obtain benefits following the requests of the confederates (Persson et al., 2013; Salahshour, 2019). The right thing was done when the confederate did it (to report the actual words), but when proposed cheating, the participant also did it (Rothstein, 2000). Remembering that the profits obtained were merged with the money contributed by each participant, in a hypothetical situation, there was the possibility that the common fund would disappear if all the pairs cheated and reported more words in all the rounds, thus fulfilling the called tragedy of the commons (Hardin, 1968).

The proposed experimental model could be used to analyze other variables. In this sense, social norms were analyzed in the second study.

Study 2: effects of descriptive norms on cooperation

This study extends the results of Study 1 by manipulating descriptive norms, a variable whose effects on cooperation and corruption are known.

Miller and Prentice (2016) define social norms as the tendency to behave as most people do; Young (2015) defines it as "unwritten codes and informal understandings that define what we expect of other people and what they expect of us" (p. 360). Cialdini et al. (1990) show that a person's behavior can be influenced by *descriptive norms*, which are observed behaviors that people commonly perform. These norms provide a quick response guide; thus, people can act as other people do in the same situation (Cialdini et al., 1991).

Thus, chaotic environments are perceived as clues of an implicit tendency to transgress the rules of order without receiving sanctions since other people have done it before (Keizer et al., 2008). However, the descriptive norm can only influence other people's behavior if it is seen as a focus of attention for others (Cialdini et al., 1991).

Fehr and Fischbacher (2004) point out that social norms determine cooperation between individuals; there will be cooperation if the other individuals cooperate. Hallsworth et al. (2017) noted that the payment of taxes increases when messages that relate to social norms are presented (e.g., "paying taxes means that we all gain from vital public services like the National Health Service, roads and schools").

Köbis et al. (2019) found in a field experiment that there was a lower bribery descriptive norm and fewer bribes in a game where participants were exposed to posters with anti-bribery messages. Similarly, Abbink et al. (2018) showed that participants offered twice as many bribes in a game when they knew they were interacting with a person whose companions were mostly corrupt, contrary to when they knew that most were honest.

The present study aims to evaluate the effects of the descriptive dishonesty norm in cooperation for personal gain. The guiding question in this study is, to what extent can a common standard of dishonesty increase cooperation for corruption? It is hypothesized that higher corruption behavior will be observed in the descriptive dishonesty norm than in the control condition.

Method

Participants

72 people participated, 56.9% women and 40.28% men, with a mean age of 21.95 (SD = 7.64). 8.57% had incomplete high school or lower studies, 50% completed high school, 25.71% had incomplete university studies, and 15.71% completed university studies. Participants were randomized; there were 36 participants in the control condition and 36 in the experimental condition. Of the 96 people invited, nine people declined the invitation, 11 agreed to participate but did not continue with the communication, and four people were excluded from the analysis for guessing at the study objectives.

Procedure

Adjustments were made to the experimental conditions of Study 1. There was only one group chat on the Hangouts platform in this study with two participants and six confederates. After each round, they were asked to report in the group chat how many words formed. In the control condition, six confederates, who supposedly played as three pairs, reported the words they had formed without exceeding the goal of five words. In the experimental condition (dishonesty norm), the six confederates reported forming more words than they did, exceeding the goal in two rounds. They reported on the public chat that they did it to earn more money and recover from the losses of the previous rounds. Thus, the participants were exposed to a descriptive norm of violating the honesty norm.

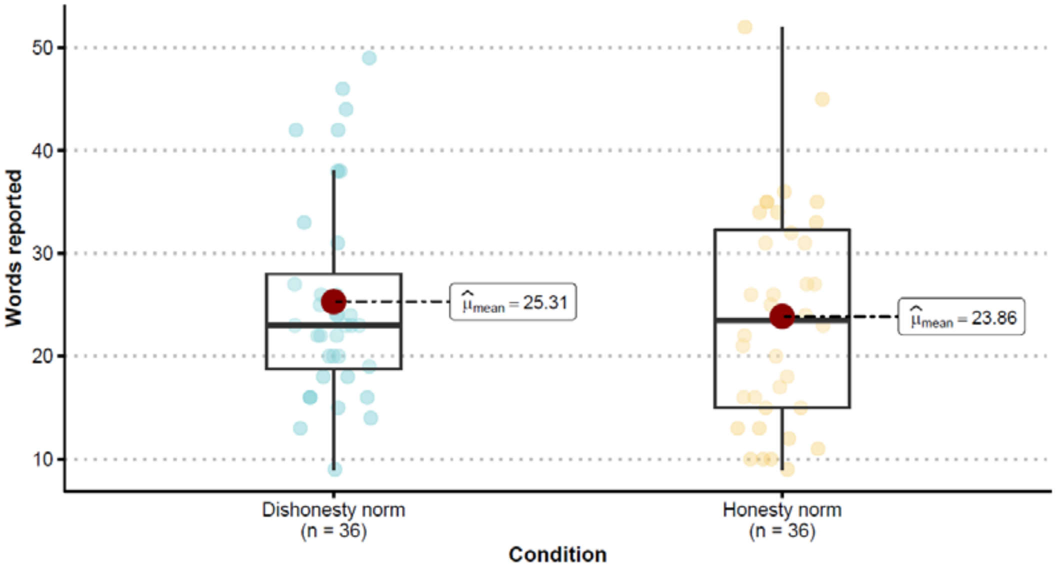
Two participants played as a pair in each experimental session in this case. They had the dilemma of whether to follow the other teams and report that they had exceeded the goal or to report the number of words they really formed. Everything else was done in the same way as in the procedure described in Study 1.

Results

In this study, 25% of the participants reported exceeding the limit in the condition with the honesty norm compared to 27.8% who exceeded it in the condition with the dishonesty norm, although the difference is not statistically significant and the effect size is very small, $\chi^2(1) = 0$, $p = 1$, 1-β = .79, V = .0, 95% CI [0, 1]. The average reported words were 25.31 (SD = 10.02) in the dishonesty condition and 23.86 (SD = 10.65) in the honesty condition (Figure 3), being this difference not statistically significant, $W = 694$, p =. 608, 1-β = .61, r_b = 0.07, 95% CI [-.19, .33].

Figure 3

Words reported by participants according to the condition in which they played (Study 2)

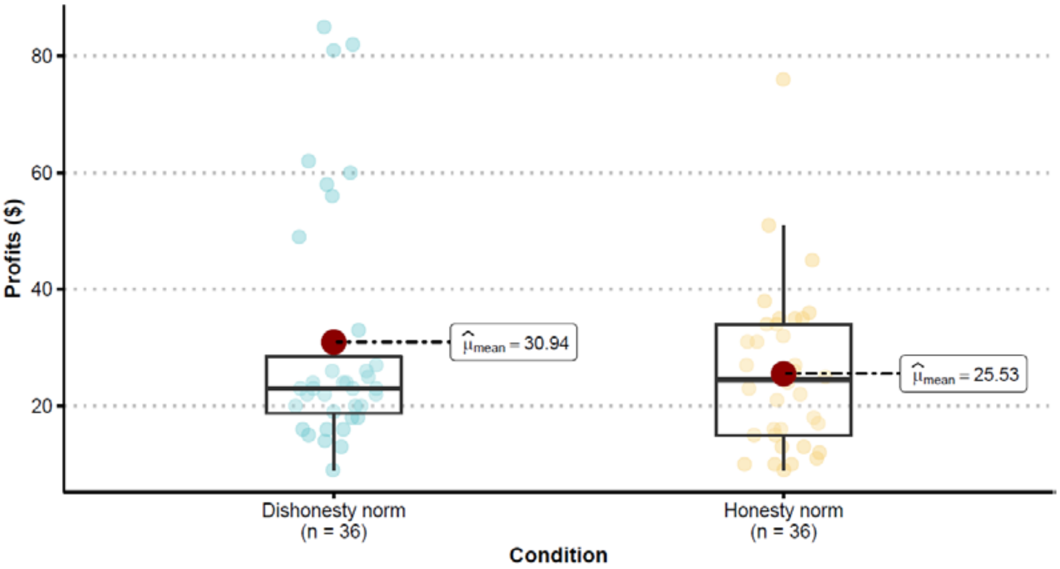


Note. Average of words reported, quartiles, and outliers in each condition

Neither were significant the differences found in the average profits for the conditions of dishonesty (\$30.94, SD = 25.79) and honesty (\$25.53, SD = 13.77), $W = 694$, $p = .608$, 1-β = .51, r_b =. 07, 95% CI [-.19, .33], as seen in Figure 4.

Figure 4

Average profits of participants according to the condition in which they played (Study 2)



Note. Average profits, quartiles, and outliers in each condition

Discussion

This study found no significant differences in the reports of exceeding the goal, the average number of words reported, or the profits received in both conditions, rejecting the hypothesis.

A first explanation of what happened is that, for a norm to affect the behavior of group members, it must be seen as a focus of attention (Cialdini et al., 1991). The participants had to focus on the other pairs reporting word counts that exceeded the limit. In the results of Cialdini et al. (1991), it is observed that those who were exposed to a model that threw garbage in a dirty environment threw more trash out of the dumpster. Still, those who saw the same model in a space where no one else had thrown garbage before followed the example of the others and not the model, not littering. It is possible that the environment generated by the experimental procedure, being a university project dedicated to science, gave the impression of a *clean environment*, where most people had to follow the rules, and only some transgressed them, but not for that, they represented a good role model.

A descriptive norm (frequency of behaviors) was used in this work. Still, an injunctive norm was not used, which is defined as those behaviors that are approved or disapproved by people explicitly (Cialdini et al., 1990). It is likely that the participants just disagreed with cheating because they did not want to transgress the honesty norm, they did not want to cause harm to third parties, or they feared some punishment for cheating. Brauer and Chaurand (2009) already indicated that injunctive norms were strongly related to social control, which could have happened in this study. Through various studies, Eriksson et al. (2015) point out that injunctive norms of disapproval can arise when conduct is considered unusual (descriptive norm). It may have been unusual to observe six confederates cheating so openly rather than privately, as usual in corrupt acts.

Study 3: social sanctions on cooperation for corruption

Fehr and Gächter (2000) found that cooperators punished non-cooperators (free riders) even though this implied a cost and did not reflect immediate gains for them. *Altruistic punishment* is the act in which people punish free riders even if this results in a cost for them and does not produce immediate material gains (Fehr & Gächter, 2002). Egas and Riedl (2008) found that cooperation is only maintained if there are optimal conditions to do the altruistic punishment, as its high impact and low cost.

Bond (2019) observed in experimental procedures of altruistic punishment that participants modify their behavior depending on the level of cooperation of others. Likewise, when the cost is low and the impact is high, the effects on the cooperation last longer and spread among more people in the network. Zhang et al. (2017) simulated free-rider punishment strategies proposing a situation where the cost of punishment is shared by cooperators, finding that there is greater cooperation when free-riders are punished and a decline in cooperation when there is no punishment.

However, other studies indicate that, although people punish people who offend them directly or people close to them, they are very unwilling to punish those who offend strangers, although they do show a certain level of anger about that situation (Pedersen et al., 2018). This willingness to punish people who offend strangers is, in fact, almost null in not experimental situations (Pedersen et al., 2020).

The research question guiding this study is: To what extent does the mechanism that allows altruistic punishment or the anonymous demand for transparency affect cooperation in corruption acts, compared to a scenario where punishment and transparency are impossible?

As a hypothesis, less cooperation to transgress the game rules for benefit increasing (corruption) is expected in the punishment and transparency conditions than in the control condition. If the anonymous demand for transparency has a lower social cost than the altruistic punishment previously used in the studies by Fehr and Gächter (2000, 2002), a greater willingness to participate against corruption is expected in the transparency condition than in the punishment condition.

Method

Participants

There was an initial sample of 133 participants, of whom 10 abandoned the procedure. Three were discarded from the final sample for guessing at the study objectives or not adequately understanding the procedures. Finally, data from 120 participants were analyzed, 50.42% were women, and 48.74% were men; the ages were between 18-70 years (*M* = 24.4, *SD* = 7.28). Regarding the level of studies, 58.8% had incomplete professional studies, 26.9% completed professional studies, and 9.2% had high school or lower studies. All were residents of the metropolitan area of Léon, Guanajuato.

Instruments

Excel registration forms were filled out online from Google platforms to enter the game. The results of the tasks assigned in each experimental round were recorded in these formats. In addition, there was a sociodemographic section where the participants answered about their age, educational level, and sex.

Procedure

It was explained to them that they would play in teams of two people, connected online with eight more players plus the experimenter; but only one player was real, and the others were confederates. The game was played on a Google spreadsheet, which was accessed with a link shared in the group chat. On the sheet, there was a table with columns of the pair and player number, along with the instructions for the game.

Excel registration forms were filled out online from Google platforms to enter the game. The results of the tasks assigned in each experimental round were recorded in these formats. The experimental session consisted of 10 rounds in which participants had to form words with letters provided. In each round, the letters were written in a table column.

The earnings of each pair appeared at the bottom of the table, added automatically as they reported the number of words. Both players earned $1 for each word they formed, which was kept in a common fund that would be shared among all players finishing all rounds. For each round, a goal of 5 words was set. If the pair exceeded that goal, the players could keep the profits in that round in a private fund, which would not be shared with the other teams, and take $1 from the common fund. With this payment scheme, players had incentives to report more words, but if many did, they would deplete the common pool for everyone, simulating the tragedy of the commons.

Participants had a Confederate teammate, who during rounds six and 10 of the procedure reported false higher data to have more profits in the private fund (corruption). These profits only would be received if the participant also reported having exceeded the word goal, signaling cooperation. Additionally, pair two reported in rounds two, three, six, nine, and 10 that they had exceeded the word limit, registering between six and eight words for each player in that pair, so their profits were higher than the other teams. This stimulus made it possible to analyze whether the participants decided, depending on the experimental condition, to punish or request that the results of that pair be investigated (transparency).

Participants were randomly assigned to one of three conditions. In the punishment condition, participants could punish other players if they believed they were cheating, paying $1 for each punishment, and subtracting the punished team's profits in that round. Punishment was openly requested in a group chat so that everyone could see it. On the condition of transparency, the participants could pay $1 for asking in the experimenter's private chat to investigate a team if they considered that it was cheating. Supposedly, if the reported team was found to be cheating, they were penalized with the profits from the round in which they were reported. Participants were informed that neither the amounts paid for transparency or punishment nor the profits taken from those allegedly cheating, would be returned to the common fund, making it clear that they would not have direct gains from these decisions. In the control condition, the participants could not request punishment or transparency as in the other conditions.

At the end of the ten rounds, the post-experimental interview was carried out. The common fund was distributed among all the participants, and ways were agreed to send them the corresponding earnings, indicating that they could also keep their initial contribution. They were informed about the experimental manipulations of the procedure, its objectives, and the contact details with the research team were reiterated.

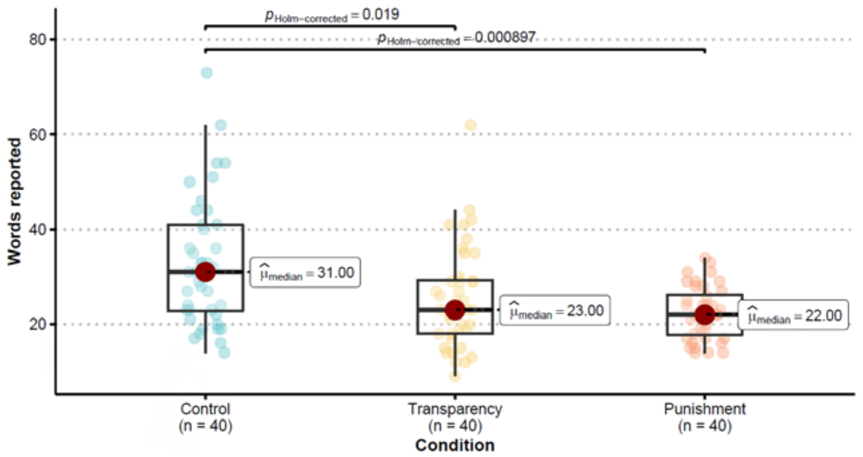
Results

In the first analysis, the number of times it was reported to exceed the goal in each condition was compared through a contingency table. As expected, in the control condition there were a greater number of reports of exceeding the goal, with 47.5% of cases, compared to 20% in the transparency condition and only 7.5% in the punishment condition, differences that are statistically significant with medium effect size, $\chi^2(2) = 17.86$, $p < .001$, $1\text{-}\beta = .43$, V = .36, CI 95% [.18, 1].

In addition, Kruskal-Wallys's test was used to analyze whether there were differences in the words reported in the experimental conditions. Again, the participants in the control condition reported a higher number of words (M = 32.83, SD = 13.66, *Mdn* = 31) compared to the conditions of transparency (M = 25.50, SD = 10.88, *Mdn* = 23) and punishment (M = 22.45, SD = 5.51, *Mdn* = 22), differences that are statistically significant with medium effect size, $H(2) = 13.90$, $p < .001$, $1\text{-}\beta = .62$, $\varepsilon^2 = .12$, CI 95% [.04, 1], as seen in Figure 5. Post hoc analysis with Dunn's test indicates that differences are generated between the control condition with the transparency condition, $p = .018$, and with the punishment condition, $p < .001$, without being significant between the transparency and punishment conditions, $p = .307$. The post hoc tests were performed with the *rstatix* package (Kassambara, 2021).

Figure 5

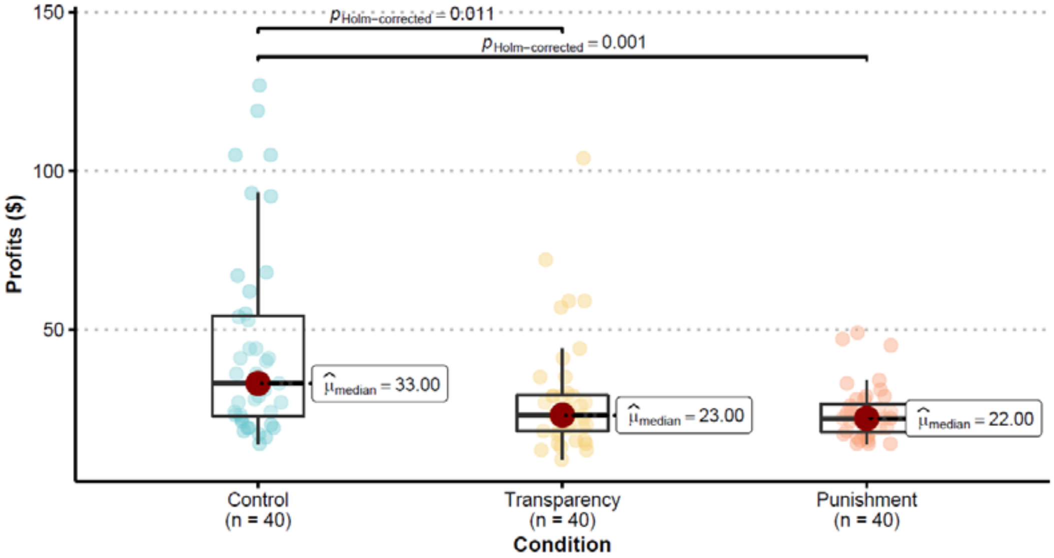
Words reported by participants according to the condition in which they played (Study 3)



Note. Medians of reported words, quartiles, and outliers are shown. Pairwise comparisons were made with Dunn's test with Holm's fit; only significant comparisons are shown

Once again, in the control condition the profits were higher ($M = 52.4$, $SD = 43.50$, $Mdn = 33$) compared to the transparency condition ($M = 30.3$, $SD = 20.69$, $Mdn = 23$) and the punishment condition ($M = 23.8$, $SD = 8.46$, $Mdn = 22$), resulting in significant differences with medium effect size, $H(2) = 14.01$, $p < .001$, $1\text{-}\beta = .69$, $\varepsilon^2 = .12$, CI 95% [.04, 1] as shown in Figure 6. Post hoc analyses with Dunn's test indicate that the differences are generated between the control condition with the transparency condition, $p = .011$, and with the punishment condition, $p = .001$, without being significant between the transparency and punishment conditions, $p = .433$.

Figure 6
Profits of participants according to the condition in which they played (Study 3)



Note. Medians of profits, quartiles, and outliers are shown. Pairwise comparisons were made with Dunn's test with Holm's fit; only significant comparisons are shown

Regarding requests for punishment or transparency, only 4 (10%) participants in the transparency condition requested a total of 14 reports. In the punishment condition, 6 (15%) participants requested a total of 20 punishments from other couples. The percentages of those who requested reports or punishments in each condition do not differ significantly from each other, $\chi^2(1) = 0.11$, $p = .735$, $1\text{-}\beta = .73$, V = .0, CI 95% [.0, 1].

Discussion

As in the studies by Fehr and Gächter (2000, 2002), implementing the possibility of punishing others turned out to be an effective measure to keep the transgression of norms to the detriment of the majority. Previous computational simulations show that sanctions against the violation of norms are more effective when they are performed by the majority than when they are concentrated on a few agents (Chen et al., 2014), although the success of these measures to preserve cooperative ties in a social system largely depend on whether sanctions are sufficiently costly (Zhang et al., 2017).

Therefore, for altruistic punishment to reduce the transgression of norms, a significant proportion of the population must forcefully punish the transgressors. However, studies show little willingness in natural conditions to punish others who have not directly affected us (Pedersen et al., 2020). Even in laboratory conditions, people prefer to avoid the responsibility of sanctioning others when they are offered the opportunity, even if they had committed to doing so (Kriss et al., 2016). This low willingness to punish others for their acts of corruption can be especially low in countries with chronic corruption problems, where even beliefs that justify corruption are established in the culture (Cruz et al., 2020). Furthermore, even with the potential for altruistic punishment to contain rule transgression, the solution cannot simply be to promote or support altruistic punishment without considering the possible consequences for the punishers. For example, Front Line Defenders (2021) reports that in 2020 at least 331 people were killed in the world for their work as human rights defenders, besides those who are threatened or persecuted even by the governments of their own countries.

Bauhr et al. (2020) review shows that transparency mechanisms effectively combat corruption. Even the governments of highly corrupt countries tend to adopt transparency and anti-corruption policies, especially in legitimacy crises, corruption scandals, and high political competition. They often adopt these policies to signal change or honesty, hoping to be in control of their implementation to maintain their impunity. But this is usually a wrong calculation. The pressures of the electorate and other internal or external political forces can promote the effective implementation of transparency and the fight against corruption (Schnell, 2018).

This study shows that a system that makes it possible to request access to transparency easily can be as dissuasive to corruption as altruistic punishment, also involving fewer risks for citizens. To delimit the actual scope of these tools, it would be necessary to assess aspects as the regulatory burden associated with requests for transparency in different nations and the willingness and capacities of citizens to make these requests.

We consider that the experimental procedure presented is a contribution that can be easily adapted in future studies to other nearby topics, for example, for the analysis of strong reciprocity (Gintis et al., 2008), where, in addition to sanctioning free-riders (strong negative reciprocity), cooperation with those who respect the rules would be privileged, even though cooperating with corrupt people could bring higher personal benefits (strong positive reciprocity).

General conclusions

We consider that the experimental model successfully represents cooperation, the formation of a common fund, the transgression of the rule of honesty, and damage to third parties in corruption, in a relatively inexpensive procedure without problems of experimental death due to abandonment or poor understanding of the rules.

Results show evidence of how public or private sanctions can be a viable strategy to maintain cooperation to prevent corruption. However, for these strategies to work, sanctions must be applied effectively, where people have the certainty that their complaints will be dealt with effectivity, without allowing impunity.

Among the limitations of this work, it is possible, as in any experimental design, that the results are due to the controlled and artificial situation, and that in real scenarios they become complicated to achieve, as some authors already mentioned regarding the effectiveness of punishment (Pedersen et. al, 2020). As Julián and Bonavia (2017) identify in their theoretical review, much of the research on corruption is done through experimental methods that can present artificial situations, but at the same time constitute a valuable tool to model possible modifications to the rules or sanctions as are presented in this study. However, the results are consistent with those found previously, favoring the idea of the possibility of generalizing them and applying them to public policies (Belaus et al. 2016).

References

Abbink, K., Freidin, E., Gangadharan, L., & Moro, R. (2018). The effect of social norms on bribe offers.

The Journal of Law, Economics, and Organization, *34*(3), 457–474. https://doi.org/10.1093/jleo/ewy015

Bacharach, M. (2018). *Beyond individual choice*. Princeton University Press.

Banks, J. S., & Sobel, J. (1987). Equilibrium selection in signaling games. *Econometrica*, *55*(3), 647. https://doi.org/10.2307/1913604

Bauhr, M., Czibik, Á., de Fine Licht, J., & Fazekas, M. (2020). Lights on the shadows of public procurement: Transparency as an antidote to corruption. *Governance*, *33*(3), 495–523. https://doi.org/10.1111/gove.12432

Belaus, A., Reyna, C., & Freidin, E. (2016). Medición y manipulación de normas sociales en juegos experimentales de corrupción. *Cuadernos de Economía*, *35*(68), 353–377. https://doi.org/10.15446/cuad.econ.v35n68.44395

Ben-Shachar, M.S., Lüdecke, D. & Makowski, D. (2020). "Effect size: Estimation of Effect Size Indices and Standardized Parameters." *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Bond, R. M. (2019). Low-cost, high-impact altruistic punishment promotes cooperation cascades in human social networks. *Scientific Reports*, *9*(1), 2061. https://doi.org/10.1038/s41598-018-38323-7

Bowles, S., & Gintis, H. (2011). *A cooperative species. Human reciprocity and its evolution*. New Jersey: Princeton University Press.

Brauer, M., & Chaurand, N. (2009). Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people's reactions to uncivil behaviors. *European Journal of Social Psychology*, *40*(3), 490–499. https://doi.org/10.1002/ejsp.640

Chen, X., Szolnoki, A., & Perc, M. (2014). Probabilistic sharing solves the problem of costly punishment. *New Journal of Physics*, *16*. https://doi.org/10.1088/1367-2630/16/8/083016

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 201–234). Academic Press. https://doi.org/10.1016/s0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Cruz, C. E., Correa, F. E., García y Barragán, L. F., & Contreras, C. C. (2020). Las creencias que justifican la corrupción disminuyen la disposición y el apoyo percibido para combatirla. *Revista Latinoamericana de Psicología*, *52*, 235–242. https://doi.org/10.14349/rlp.2020.v52.23

Dang, V. C., Nguyen, Q. K., & Tran, X. H. (2022). Corruption, institutional quality and shadow economy in Asian countries. *Applied Economics Letters*, 1–6. https://doi.org/10.1080/13504851.2022.2118959

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1637), 871–878. https://doi.org/10.1098/rspb.2007.1558

Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, *129*, 59–69. https://doi.org/10.1016/j.obhdp.2014.09.011

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

Fierro, H. F., Flores, J., Ayllón, S. L., & Valadés, D. (2011). Encuesta nacional de cultura de la legalidad. [Available online at: https://constitucion1917.gob.mx/work/

models/Constitucion1917/Resource/1073/1/images/EncuestaConstitucion_UNAM_2011.pdf].

Fierro, H., Flores, J., & Valadés, D. (2017). *Los mexicanos y su Constitución. Tercera Encuesta Nacional de Cultura Constitucional. Los mexicanos vistos por sí mismos. Los grandes temas nacionales* (Primera, Vol. 3). México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas.

Fišar, M., Kubák, M., Špalek, J., & Tremewan, J. (2016). Gender differences in beliefs and actions in a framed corruption experiment. *Journal of Behavioral and Experimental Economics*, *63*(1), 69–82. https://doi.org/10.1016/j.socec.2016.05.004

Front Line Defenders (2021). Front line defenders global analysis 2020. [Available online at: https://www.frontlinedefenders.org/sites/default/files/fld_global_analysis_2020.pdf].

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research*, *21*(2), 241–253. https://doi.org/10.1007/s11211-008-0067-y

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31. https://doi.org/10.1016/j.jpubeco.2017.02.003

Hardin, G. (1968). The tragedy of the commons. *Science*. https://doi.org/10.1126/science.162.3859.1243

Heinz, M., & Schumacher, H. (2017). Signaling cooperation. *European Economic Review*, *98*, 199–216. https://doi.org/10.1016/j.euroecorev.2017.06.017

INEGI, (2015). Segunda encuesta nacional de calidad e impacto gubernamental. [Available online at: https://www.inegi.org.mx/programas/encig/2015/].

Julián, M., & Bonavia, T. (2017). Aproximaciones psicosociales a la corrupción: una revisión teórica. *Revista Colombiana de Psicologia*, *26*(2), 231–243. https://doi.org/10.15446/rcp.v26n2.59353

Kassambara, A. (2021). *rstatix: Pipe- Friendly framework for basic statistical tests* (version 0.7.0). R package. https://cran.r-project.org/web/packages/rstatix/index.html

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, *322*(5908), 1681–1685. https://doi.org/10.1126/science.1161405

Köbis, N. C., Troost, M., Brandt, C. O., & Soraperra, I. (2019). Social norms of corruption in the field: Social nudges on posters can help to reduce bribery. *Behavioral Public Policy*, 1–28. https://doi.org/10.1017/bpp.2019.37

Kriss, P. H., Weber, R. A., & Xiao, E. (2016). Turning a blind eye, but not the other cheek: On the robustness of costly punishment. *Journal of Economic Behavior and Organization*, *128*, 159–177. https://doi.org/10.1016/j.jebo.2016.05.017

Legorreta, A., & de Mola, G. R. L. (Eds.). (2016). *Corrupcionario mexicano*. Grijalbo.

Marquette, H., & Peiffer, C. (2015). *Collective action and systemic corruption*. (Paper presented at the ECPR Joint Sessions of Workshops). [Available online at: https://baselgovernance.org/publications/collective-action-and-systemic-corruption].

Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology*, *67*, 339–361. https://doi.org/10.1146/annurev-psych-010814-015013

Mungiu-Pippidi, A. (2013). Controlling corruption through collective action. *Journal of Democracy*, *24*(1), 101–115. https://doi.org/10.1353/jod.2013.0020

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755

Ostrom, E. (2003). How types of goods and property rights jointly affect collective action. *Journal of Theoretical Politics*, *15*(3), 239–270. https://doi.org/10.1177/0951692803015003002

Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, *100*(3), 641–672. https://doi.org/10.1257/aer.100.3.641

Ostrom, E. (2011). *El gobierno de los bienes comunes. La evolución de las instituciones de acción colectiva*. Fondo de Cultura Económica.

Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software*, *6*(61), 3167, https://doi.org/10.21105/joss.03167

Pedersen, E. J., McAuliffe, W. H., & McCullough, M. E. (2018). The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically. *Journal of Experimental Psychology: General*, *147*(4), 514-544. https://doi.org/10.1037/xge0000410

Pedersen, E. J., McAuliffe, W. H. B., Shah, Y., Tanaka, H., Ohtsubo, Y., & McCullough, M. E. (2020). When and why do third parties punish outside of the lab? A cross-cultural recall study. *Social Psychological and Personality Science*, *11*(6), 846–853. https://doi.org/10.1177/1948550619884565

Persson, A., Rothstein, B., & Teorell, J. (2013). Why anticorruption reforms fail-systemic corruption as a Collective Action problem. *Governance*, *26*(3), 449–471. https://doi.org/10.1111/j.1468-0491.2012.01604.x

Piccolo, S., & Immordino, G. (2017). Organized crime, insider information, and optimal leniency. *Economic Journal*, *127*(606), 2504–2524. https://doi.org/10.1111/ecoj.12382

R Core Team. (2021). *R: A language and environment for statistical computing* (version 4.1.0). R Foundation for Statistical Computing. https://www.r-project.org/

Rivas, M. F. (2013). An experiment on corruption and gender. *Bulletin of Economic Research*, *65*(1), 10–42. https://doi.org/10.1111/j.1467-8586.2012.00450.x

Rothstein, B. (2000). Trust, social dilemmas and collective memories. *Journal of Theoretical Politics*, *12*(4), 477–501. https://doi.org/10.1177/0951692800012004007

Salahshour, M. (2019). Evolution of costly signaling and partial cooperation. Scientific Reports, 9(1), 1–7. https://doi.org/10.1038/s41598-019-45272-2

Schnell, S. (2018). Cheap talk or incredible commitment? (Mis)calculating transparency and anti-corruption. Governance, 31(3), 415–430. https://doi.org/10.1111/gove.12298

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. The American Economic Review, 92(3), 434–459. https://doi.org/10.1257/00028280260136200

Spyromitros, E., & Panagiotidis, M. (2022). The impact of corruption on economic growth in developing countries and a comparative analysis of corruption measurement indicators. *Cogent Economics and Finance*, *10*(1), 2129368. https://doi.org/10.1080/23322039.2022.2129368

Sutherland, E. H. (1940). White-Collar Criminality. *American Sociological Review*, *5*(1), 1–12. https://doi.org/10.2307/2083937

Transparency International. (2016). Corruption perceptions index 2015. [Available online at: https://www.transparency.org/whatwedo/publication/cpi_2015].

Transparency International (2019). Anti-Corruption Glossary. [Available online at: https://www.transparency.org/glossary/term/corruption].

Transparency International. (2021). Corruption perceptions index 2020. [Available online at: https://www.transparency.org/en/cpi/2020/index/nzl].

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of Public Economics*, *76*(3), 399–457. https://doi.org/10.1016/S0047-2727(99)00092-4

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651–10656. https://doi.org/10.1073/pnas.1423035112

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics*, 7(1), 359–387. https://doi.org/10.1146/annurev-economics-080614-115322

Zhang, C., Zhu, Y., Chen, Z., & Zhang, J. (2017). Punishment in the form of shared cost promotes altruism in the cooperative dilemma games. *Journal of theoretical biology*, *420*, 128-134. http://dx.doi.org/10.1016/j.jtbi.2017.03.006

Appendix A

Provided letters and possible words in each game round

Round	Letters	Possible words							
Test 1	Q A C B O A	ABACO	CAOBA	ABOCA	ACABO	BACA	BOCA	CABO	COBA
Test 2	O P S A	SOPA	SAPO	POSA	PASO	PASÓ	ASPO	ASPÓ	OPAS
1	E H A V T G	VEGA	GETA	VAHE	VETA	VATE			
2	V G A A P H	PAVA	HAGA	PAGA	VAGA	VAHA			
3	Q O T O R	ROTÓ	ROTO	TORO	OTRO	ORTO			
4	P H E Q T A	PATÉ	PATE	PETA	TAPE	TAPÉ			
5	A X A M S A	AMAS	ASMA	SAMA	MASA	AMASA			
6	X T H A J E	JATE	TEJA	TAJÉ	TAJE	JETA			
7	T T Q O R U	TUTOR	TOUR	TUTO	RUTO	RUTÓ			
8	L I Q E F C	FIEL	FICE	FICÉ	FILE	FILÉ			
9	S O Q L W O	SOLO	OSLO	LOSO	LOSÓ	SÓLO			
10	O A P Q L A	APOLA	LAPO	PALO	PALA	LAPA			

Note: The possible words were obtained in a word generator considering the instructions of the game, that is, words of at least four letters in spanish. The resulting words were verified in the dictionary of the Royal Spanish Academy.