Michael Muthukrishna, Patrick Francois, Shayan Pourahmadi and Joseph Henrich

Corrupting cooperation and how anti-corruption strategies may backfire

Article (Accepted version) (Refereed)

Original citation:
Muthukrishna, Michael and Francois, Patrick and Pourahmadi, Shayan and Henrich, Joseph (2017) *Corrupting cooperation and how anti-corruption strategies may backfire.* Nature Human Behaviour, 1 (7). 0138. ISSN 2397-3374

DOI: 10.1038/s41562-017-0138

Corrupting Cooperation and How Anti-Corruption Strategies May Backfire

Michael Muthukrishna[*1,6], Patrick Francois[2,3], Shayan Pourahmadi[4] & Joseph Henrich[2,3,5,6]

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science, London WC2A 2AE, UK
[2]Vancouver School of Economics, University of British Columbia, Vancouver, BC V6T 1L4, Canada
[3]Canadian Institute for Advanced Research
[4]Department of Economics, Columbia University, New York, NY 10027, USA
[5]Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[6]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

Corresponding Author:
Michael Muthukrishna
Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK
+442078523612
m.muthukrishna@lse.ac.uk

Corrupting Cooperation and How Anti-Corruption Strategies May Backfire

Understanding how humans sustain cooperation in large, anonymous societies remains a central question of both theoretical and practical importance. In the laboratory, experimental-behavioral research using tools like Public Goods Games suggests that cooperation can be sustained by institutional punishment—analogous to governments, police forces, and other institutions that sanction free-riders on behalf of individuals in large societies[1-3]. In the real world, however, corruption can undermine the effectiveness of these institutions[4-8]. Levels of corruption correlate with institutional, economic, and cultural factors, but the causal directions of these relationships are difficult to determine[5,6,8-10]. Here we experimentally model corruption, by introducing the possibility of bribery. We investigate the effect of structural factors (leader's punitive power and economic potential), anti-corruption strategies (transparency and leader investment in the public good), and cultural background. The results reveal that: (a) corruption possibilities cause a large (25%) decrease in public good provisioning; (b) empowering leaders *decreases* cooperative contributions (in direct opposition to typical institutional punishment results); (c) growing up in a more corrupt society predicts more acceptance of bribes; and, (d) anti-corruption strategies are effective under some conditions, but can further decrease public good provisioning when leaders are weak and economic potential is poor. These results suggest that a more nuanced approach to corruption is needed, and that proposed panaceas, such as transparency, may actually be harmful in some contexts.

Cooperation, particularly large-scale anonymous cooperation, remains an important puzzle to both evolutionary and social scientists, with real world social and economic implications. One method for sustaining cooperation that has received considerable attention involves costly punishment[11-13], whereby individuals pay a cost to punish free riders who fail to contribute to the public good. While cross-cultural evidence shows the ubiquity of costly punishment in large-scale societies (though not in small-scale societies), there is some variability in both the motivation to punish free riders and the tendency to punish cooperators (i.e. some societies display significant levels of anti-social punishment, the punishment of cooperators)[14-16].

Research on the role of peer punishment in sustaining cooperation reveals two major challenges: (a) the second order free-rider problem in which individuals defect on the job of punishing and thereby increase their payoffs[17,18], and (b) the problem of counter-punishment—punishment as revenge for previously being punished[12,19]. Institutional, or pool, punishment resolves these problems by designating one individual as a “Leader” who can extract taxes and punish free-riders on behalf of other players[2]. Institutional punishment reduces the problems of both second-order free-riding and counter-punishment, and may thus be important in explaining the emergence and maintenance of large-scale cooperation[3]. Moreover, recent empirical research shows that participants (at least WEIRD participants[20]) prefer institutional punishment to peer punishment[1,21].

Institutional punishment, as typically modelled in Public Goods Games (PGGs), serve to incentivize player choices when contributing to the public pool, and work by constraining leader choices to either punishing players or doing nothing. In the real-world, however, channels such as bribery, nepotism, and lobbying allow individuals (or corporations) to avoid contributing to the public pool (e.g., evading taxes) as well as avoid being punished (e.g., by paying a bribe instead)—real leaders and institutions are corruptible.

Corruption is widespread, unevenly distributed, and costly. The World Bank estimates that worldwide, US$1 trillion is paid in bribes alone[7]. However, levels of corruption vary considerably. In Kenya, estimates suggest that 8 out of 10 interactions with public officials require a bribe and that the average urban Kenyan pays a bribe 16 times a month[22]. In contrast, the average Dane may never pay a bribe in their lifetime; Denmark has the lowest level of corruption based on the Corruption Perceptions Index[23]. The predicted costs of corruption vary from reductions in food redistribution anti-poverty programs[24] to deaths from collapsed buildings[4]. Most recently, corruption has been identified as a contributing factor to the Greek economic crisis—Greece has the highest level of corruption in the European Union (EU), with recent estimates placing it near China and Brazil[23]. Corruption in EU states, such as Greece, potentially undermine the future of the EU. But, while levels of corruption correlate with institutional, economic, and cultural factors, the causal interconnections among these factors remain difficult to disentangle[8,9,25].

To model corruption, we modified the institutional punishment PGG (IPGG from herein). In a public goods game, players are given an endowment which they can divide between themselves and a public pool. The public pool is multiplied by some amount and then divided equally among players, regardless of contribution. A cooperative dilemma is created by setting the multiplier such that it is in every player's best interest to let others contribute, contributing nothing themselves, but in the group's best interest for all players to contribute their entire endowment so that they all reap the maximum benefits of the multiplier. In the IPGG, one player is randomly selected as a leader who can punish using taxes extracted from other players. Past research has shown the effectiveness of using designated leaders as institutional punishers[1,2,21].

To introduce bribery, we modified the IPGG by giving players and leaders one additional choice, thereby creating the Bribery Game (BG). Players in addition to dividing their endowment between themselves and the public pool can also offer some of this endowment to improve the leader's payoff (i.e., effectively offering a bribe, though we use neutral language). In turn, leaders have an additional exclusive choice in addition to punishing or doing nothing to players: leaders can choose to take the contribution (i.e., accept the bribe) or not. We chose to make punishing, accepting bribes, or doing nothing to each player an exclusive choice for simplicity and because past research suggests that a non-exclusive choice would reduce or remove the impact of the bribe on decision-making[10]—in reality, a bribe with no effect wouldn't last long. A new leader was selected each round to remove any reputational effects, turning the game into a series of repeated one-shot encounters. We manipulate the pool multiplier (a proxy for economic potential) and the punishment multiplier (the power of the leader to punish). In the BG, we also introduce three corruption mitigation strategies: partial transparency (revealing leader contributions), full transparency (revealing all leader behavior, including bribe taking), and leader investment (forcing leaders to contribute their endowment to the public pool). We focus on transparency and discuss leader investment, which requires further investigation, in the Supplementary Information. We ran the experiment using a Canadian Economic Subject Pool open to the public, with native-born, and first and second-generation immigrants with diverse backgrounds.

We assume players (a) bring cultural differences, shaped by their different ethnic backgrounds and cultural exposure, into the game and (b) adjust their behaviours via exposure to the experimental setting, closer to the equilibrium that maximizes payoffs in the game. We model an IPGG with a fixed tax rate to more realistically capture a world in which taxes are not directly correlated with punishment and where leaders punish without a large cost to themselves (since their own taxes are a small part of the taxes contributing to the pool punishment or institution). We then modify this game to a BG by offering players and leaders the choice to offer and accept bribes. It is easy to see that without punishment, contributions will tend toward zero—contribution levels are contingent on the strength of leaders and their tendency to punish low contributors. We predict that leaders will use taxes to punish in the IPGG, since these are not personally costly and benefit the leader's payoff by increasing the size of the public good.

Increased leader strength predicts higher contributions and more public good provisioning. In the PG, we predict that players have no incentive to offer contributions or bribes unless they are punished for not doing so. However, when bribery is an option, leaders have a greater incentive to punish people for not offering brides than for not contributing, since their share of the public good will be smaller than a bribe multiplied by every player. More power gives leaders an increased ability to impose their will, increasing the rate of bribery at the expense of the public good. Thus, in contrast to the IPGG, we predict that stronger leaders in the BG should *reduce* contributions and public good provisioning. However, if players have a preference for contributions over bribes (such as if their prior experience is a world where potential returns on the public good are higher or where anti-corruption norms are adaptive), then the incentive to punish brides over contributions is dampened. In contrast, growing up in a more corrupt society may lead to a higher preference for eliciting, offering, and accepting bribes. The full set of predictions are provided in Supplementary Information.

Cost of corruption. To examine the costs of corruption, we compared the IPGG and BG. We find that when bribery is an option, mean contributions drop by 25%. The difference between these conditions (estimated using an MCMC GLMM regression; Table S2) represents a 0.43 [CI_{95}: -0.49,-0.38] standard deviation loss (1.4 points per period; equivalent to 14% of initial endowment or $2.10 over the course of the game). Not surprisingly, when corruption can enter, it does, and cooperation deteriorates.

Causes of corruption. Having established the impact of bribery on cooperation, we examine the causes of this corruption. In Table 1 and Figure 1 we use an MCMC categorical GLMM to estimate the effect of (1) our different treatments, (2) cultural experience and (3) background on leader decisions. Leaders with a stronger punishment multiplier at their disposal ("Stronger Leader") were about twice as likely to accept bribes and about 3 times less likely to do nothing. In contrast, when accepting bribes is not an option (IPGG), these more powerful leaders were as likely to do nothing (see Leader Decisions in Supplementary Information). Thus, as expected, more power led to more corrupt behavior.

Exploring individual variation, we found that those who grew up in more corrupt countries were more willing to accept bribes. For every one standard deviation increase in players' exposure corruption scores (see Corruption Perception Scores in Supplementary Information for details on how these scores were constructed and the distribution of these scores in our sample), leaders were 1.2 times more likely to accept a bribe. In contrast, when players' parental heritage included countries with higher corruption norms (i.e. more perceived corruption), leaders were 1.5 times *less* likely to accept bribes for every standard deviation increase in corruption score, and 1.6 times more likely to do nothing (see Figure 1; Supplementary Information shows all models). In combination with other evidence[5,6,26-29], we suspect our corruption exposure scores are capturing internalized social norms related to corruption acquired while growing up in different communities. Meanwhile, our parental heritage effects, which are driven by our Canadian-born participants (e.g., second generation immigrants), may capture an internalized reaction against ethnic stereotyping—i.e., a reaction against the assumption of corrupt behavior from those of their heritage[21].

	Accept Bribe	**Punish**	**Do Nothing**
High Economic Potential	1.37 [0.65,2.21]	0.79 [0.41,1.14]	0.81 [0.29,1.40]
Strong Leader	2.14 [1.18,3.36]	1.08 [0.60,1.61]	0.29 [0.10,0.50]
Player Exposure Corruption Score	1.22 [1.01,1.44]	0.99 [0.81,1.19]	0.79 [0.63,1.02]
Player Heritage Corruption Score	0.65 [0.54,0.79]	1.17 [0.96,1.40]	1.55 [1.25,1.89]
(Intercept)	0.57 [0.05,1.54]	0.16 [0.02,0.39]	3.01 [0.12,9.50]
Obs.	1396	1396	1396
N	175	175	175
Groups	45	45	45
DIC	36.13	18.23	18.45

Table 1 | Leader Decision Each column reports the odds ratio and 95% confidence interval of the behavior in the column heading (e.g. Accept Bribe in column 1) compared to engaging in one of the other two behaviors (e.g. Punish or Do Nothing for column 1). The odds are estimated using an MCMC categorical GLMM, with the behavior coded as 1 and the other two behaviors coded as 0. The confidence intervals are Highest Posterior Density (HPD) intervals. Each model regresses the behavior in the Bribery Game (with no transparency or leader investment) on economic potential (low vs. high), leadership strength (weak vs. strong), and both player's and leader's exposure corruption score (z-score) and heritage corruption score (z-score), controlling for period, order of conditions, order of background questions, group size, age, and gender with random effects for individuals within groups. Here we report only the predictors of interest. The full model is reported in Supplementary Information.

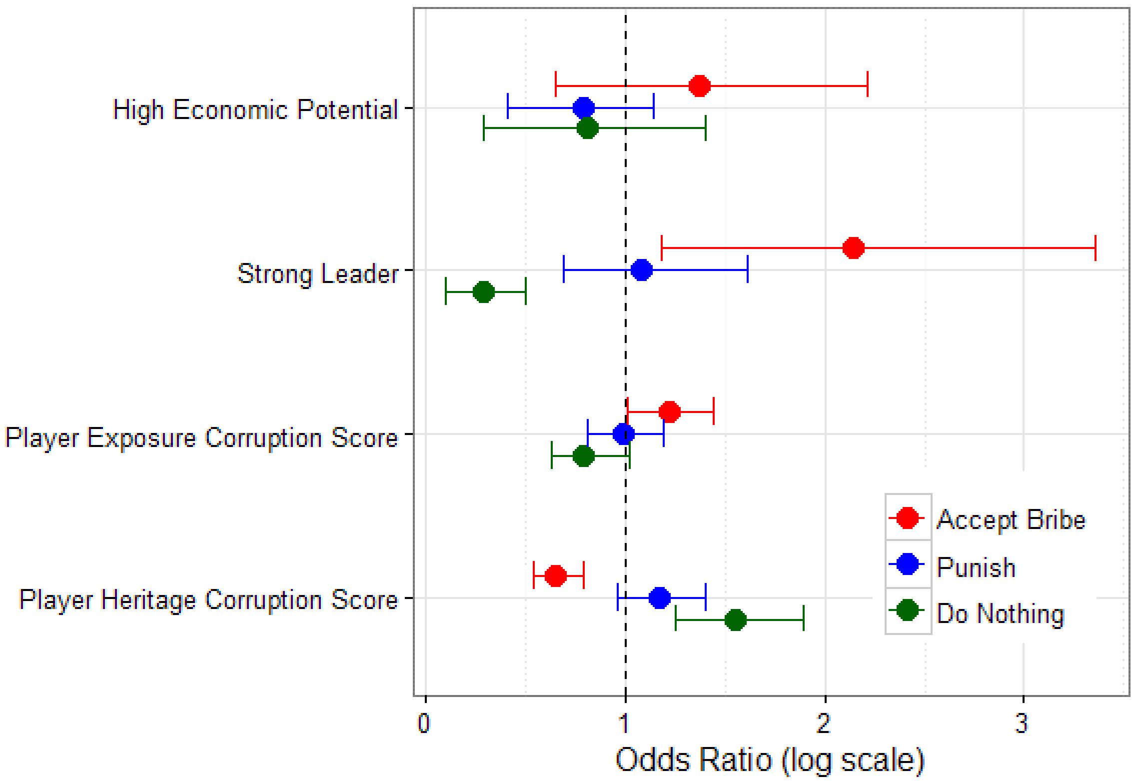

Figure 1 | Leader Decision A graphical representation of the odds ratio and 95% confidence interval of the behavior in Table 1.

Cures for corruption. Having generated corruption, we attempted to suppress it by modifying the BG using two different forms of transparency measures and by forcing leaders to invest in the public good. The first transparency approach, Partial Transparency, allowed all players to see the leader's contribution, thereby offering leaders an opportunity to establish or reveal a norm by revealing to players how much or how little leaders invested in the public pool. The second transparency approach, Full Transparency, allowed players to see all leader actions: leader contributions, the anonymized contributions and bribes from each player, and the leader's decision in each case. Leader Investment forced leaders to maximally contribute their endowment to the public good, thereby tying a large part of their payoff to the efficiency of the public good. Tying leader payoffs to the success of the public good has explicitly been used as one aspect of an anti-corruption measure in places such as Singapore, which has one of the lowest levels of corruption (based on Corruption Perception Index[23]) and the highest paid leader in the world[30]. Singaporean minister salaries are pegged at the salaries of top professionals and Singapore's Gross Domestic Product (GDP). The Leader Investment treatment is designed to be similar to linking leader payoffs to a country's GDP, but in a way that minimally deviates from the other treatment designs. This treatment, though interesting, has certain caveats in its interpretation, and requires further investigation. We report its effect and discuss these issues in more detail in the Supplementary Information.

To determine the effectiveness of these anti-corruption measures, we compared contributions in each condition to the IPGG (control) and to the BG. We regressed contributions (z-scores) on treatment, economic potential, and leader strength. Figure 2 summarizes the results of this regression and reports separate coefficients within each condition. Note that these values come from a single model and are calculated by changing reference groups (see Supplementary Information). Raw mean contribution values are graphed in Figure 3.

[INSERT FIGURE 2]

		Weak Leaders		Strong Leaders	
		Control	BG	Control	BG
Poor Economic Potential	Control	-	0.21**	-	0.52***
	Bribery Game (BG)	-0.21***	-	-0.53***	-
	BG + Partial Transparency	-0.31***	-0.10*	-0.53***	-0.01
	BG + Full Transparency	-0.20***	-0.01	-0.06	0.47***
		Control	BG	Control	BG
Rich Economic Potential	Control	-	0.39***	-	0.57***
	Bribery Game (BG)	-0.39***	-	-0.57***	-
	BG + Partial Transparency	-0.30***	0.09+	-0.44***	0.13**
	BG + Full Transparency	-0.15**	0.24***	-0.25***	0.32***

***** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$**

Figure 2 | Cures for corruption Coefficients and colors indicate the effect on contributions, on public goods provisioning. Deeper blue shading indicates greater public goods provision and darker red indicates reduced public goods. All coefficients are extracted from a single model by changing reference groups. The 4 large rectangles show the effect of each treatment with weak and strong leaders (columns) and poor and rich economic potential (rows). Within each rectangle, the columns represent the reference group treatment (Control, BG). Each row reports the coefficient of each treatment compared to this reference group. Contributions are z-scores, so coefficients represent standardized differences. The full model is reported in the SI. In all models, we account for the clustering inherent in the experimental design by including a fixed effect for the number of subjects and random effects for participants within groups. Note that in all treatments and structural contexts, the BG has lower contributions than the structurally equivalent IPGG (control). Corruption mitigation effectively increases contributions (though not to control levels) when leaders are strong or economic potential is rich. When leaders are weak and economic potential is poor, the apparent corruption mitigation strategy Full Transparency has no effect and Partial Transparency further *decrease* contributions.

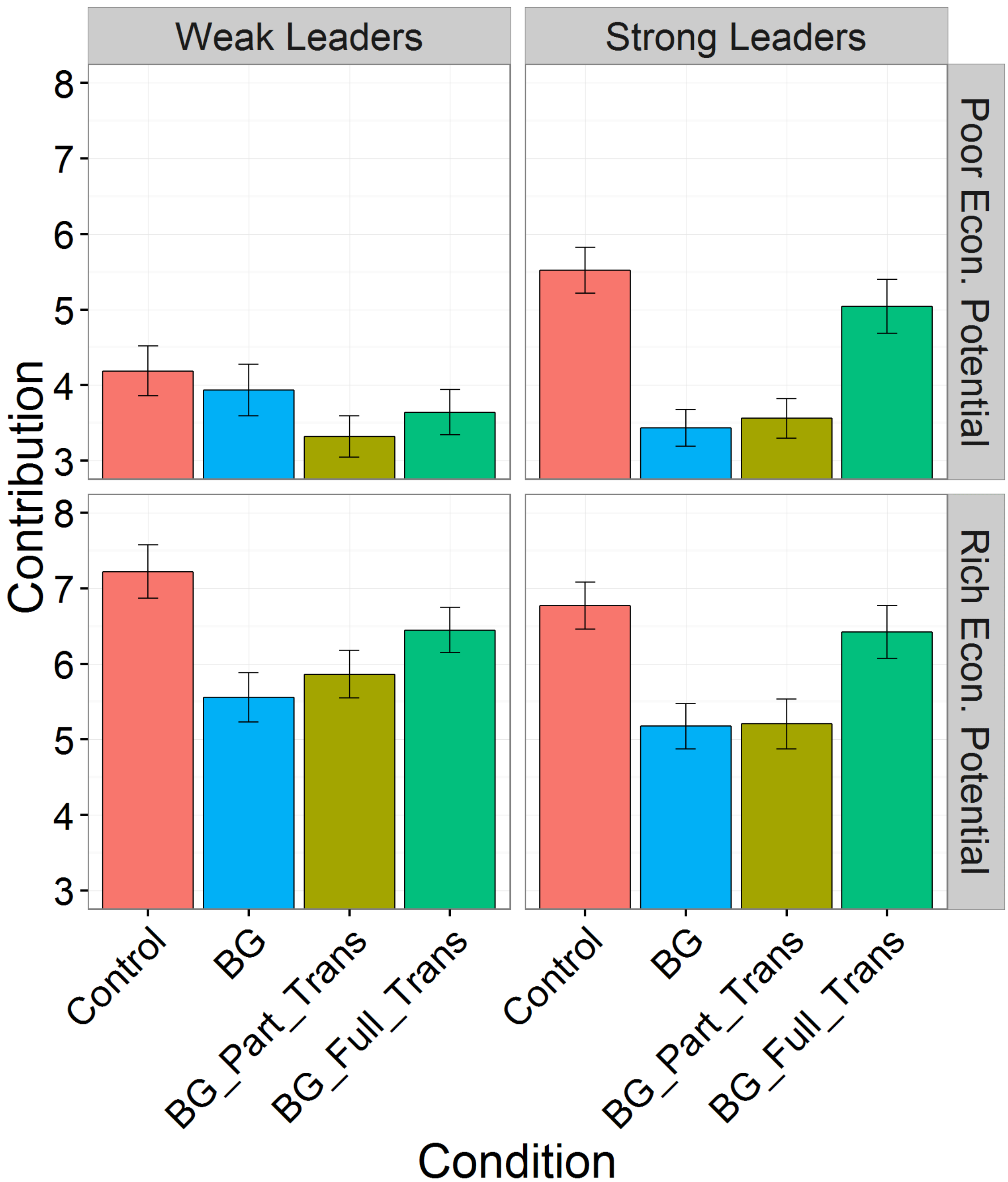

Figure 3 | Contributions by condition Raw contributions for each within-subject treatments in between-subjects structural contexts with 95% confidence intervals. These data are consistent with our theory, which predicts that more powerful leaders increase contributions in the IPGG, but decrease contributions in the BG.

Figure 2 and 3 reveal that stronger leaders are better able to increase the efficiency of public goods provisioning when economic potential is poor and bribery is not an option (red bars in top row), but when bribery is an option (blue bars), stronger leaders in poor contexts reduce the efficiency of the public good, making themselves wealthy at the expense of other players. Corruption mitigation effectively increases contributions (though not to control levels) when leaders are strong or economic potential is rich. When leaders are weak and economic potential is poor, the apparent corruption mitigation strategy Full Transparency has no effect and Partial Transparency further *decreases* contributions to levels lower than the standard BG—it led to less public good provisioning.

Although the cost of bribery was seen in all contexts, in poor economic contexts, the already low contributions were reduced even further. That is, even if powerful leaders are accepting bribes at comparable levels in both poor and rich economic contexts, the degree of corruption may not be as visible if economic potential is high. Leaders in richer economic contexts, like the United States, may accept "bribes" in the form of lobbying or campaign funding, which may indeed reduce the efficiency of the public good, but this cost isn't as obvious since economic potential is already much higher than in other nations. In contrast, in poorer economic contexts like the Democratic Republic of Congo, corruption further reduces the already low public good provisioning. Unfortunately, our results suggest that in these contexts with weak institutions and poor economic potential, efforts to mitigate corruption, such as transparency or leader investment, could backfire, further *reducing* investments in the public good. These results reflect leaders lacking the power to increase contributions through punishment and thus recouping the cost of their investment in the public good by accepting bribes. Transparency in this context, reveals a low contribution norm. Thus, the lessons in fighting corruption when institutions have the power to sustain public goods (if only corruption were reduced) and the potential for economic growth is high, may not only fail to apply when these conditions are not met, but can worsen the situation.

Our results suggest that the effect of exposure to different institutions and norms persist after moving to a new environment. This increase in corrupt behavior with direct exposure to corrupt institutions or norms is consistent with the internalization of perceived norms[5,6,26,27] and with previous empirical data showing, for example, that diplomats from high-corruption countries accumulate more unpaid parking violations[29]. However, the decreased probability of accepting bribes among those whose cultural background includes more corrupt countries, suggests that second-generation and later migrants are not as corrupt as their peers from less corrupt nations. This may represent the self-selection of immigrants from their home countries or may be a form of "identity denial"[21], whereby acculturated individuals actively avoid the stereotypes of their inherited ethnic labels. Although we have a large range of corruption scores (see Corruption Perception Scores in Supplementary Information), our sample is limited to migrants in a Canadian context and further investigation is required to determine if these cultural results generalize. Together these results suggest that corruption may be rooted in structural factors, but that internalized corruption norms may cause these behaviors to persist in a new context.

Overall, these results suggest that: (a) stronger institutions and leaders are required to sustain public goods contributions when economic potential is poor and the incentive to free-ride is high; (b) in this context, when able to, leaders will abuse their power with a noticeable economic cost. However, (c) even if economic potential is poor, if leaders are powerful, anti-corruption measures can be effective at increasing public good provisioning. Thus, efforts to mitigate corruption in poorer economic contexts must go hand-in-hand with strengthening institutions. When leaders have less punitive power, efforts such as transparency may have no effect or even decrease contributions since they reveal the rationality of low public good contributions and that most leaders do not contribute. In a rich context with powerful punitive institutions, there may be multiple equilibria that just require norms (activated in our game by transparency) to stabilize a higher payoff. In contrast, in a poor context with weak institutions, there is only one equilibria—bribe offers and low public good provisioning.

Though these experimental results begin to offer insights into the causal effect of corruption on cooperation, extending such experimental findings demands great caution. Laboratory work on the causes and cures of corruption must inform and be informed by real world investigations of corruption from around the globe. Thus, aiming only to drive future investigation, our results suggest that as economic potential grows, less government intervention is required to enforce cooperation and increased power may be misused, requiring greater anti-corruption efforts. In contrast, when economic potential is poor, strong government intervention is most effective at decreasing free-riding, as long as this intervention is paired with strategies to mitigate corruption. This may help explain why intuitions about "cures for corruption" based on experiences in rich nations do not work as well in poorer nations.

Methods

Participants. We had a total of 274 participants (166 female, mean age 20.90) drawn from an Economic Subject Pool open to the public. Participant ethnic backgrounds were as follows: 63 Euro Canadians, 158 East Asians, 17 South Asians, 36 Other Ethnicities. Participants played in groups of between 4 and 7 players. Ethical approval was obtained from the UBC Behavioural Research Ethics Board (H12-02457). Informed consent was obtained from all participants prior to beginning the study. Participants were randomly assigned to experimental groups.

Experimental Design. We used a 2 (high vs low economic potential) x 2 (weak vs strong leader power) between-subjects experimental design with 5 within-subject treatments: (institutional punishment public goods game; control (n=205), bribery game (n=222), bribery game with partial transparency (n=228), bribery game with full transparency (n=204), bribery game with leader investment(n=196)) with random allocation to all treatments. Sample size in the low economic potential, weak leader power was 71; low economic potential, strong leader power was 68; high economic potential, weak leader power was 68; and high economic potential, strong leader power was 67.

In the real world, leaders make institutional decisions based on a fixed budget to which they are one among many contributors and which has to be spent. To better model these conditions, we extracted fixed taxes for punishment, which were discarded if not used. Participants were randomly assigned to one of the 4 between-subjects treatments and 4 of the 5 within-subjects treatments.

To test possible contributing causes of corruption, we randomly assigned each group of participants to a treatment with either high or low (i) marginal per capita rate of return (0.3 vs. 0.6) as a measure of economic potential and (ii) a punishment multiplier (1 vs. 3) as a measure of the strength of the leader or institution. The marginal per capita rate of return is the expected return for every point invested in the public pool and the punishment multiplier is the number of points subtracted from a sanctioned player for every tax point spent on punishing that player.

Within-subject treatments were played in a random order with pre-recorded video instructions prior to each period. A quiz was given prior to beginning, to ensure participants knew how each treatment worked. This quiz along with the script and screenshots from the video can be found in Supplementary Information. We used a block randomization design, where participants played a minimum of 10 rounds, but the game may have ended at any point prior to the completion of 10 rounds. At 10 rounds, participants were informed which round the period ended or played further rounds until the game ended. In this way, we had 10 rounds to analyze without end game effects—participants did not know when the game would end. In order to remove reputational effects, for each round, the leader was also randomly selected. Random selection was with replacement, such that players also couldn't say that the same person couldn't be the leader for a consecutive round. As such, the experiment can be interpreted as a series of one-shot interactions. Participants were paid for 10 random rounds from across all conditions. They were paid at a rate of 15c per point, with a show up fee of $10.

Measures. We measured Age, Gender, University Degree or Occupation and Major or Industry, Prestige/Dominance, Right Wing Authoritarianism, whether they spent their entire life in Canada, where else they've lived, what suburb they grew up in, ethnic group, Religion and importance of religion, how well they speak the language of their ethnic heritage (Cultural Competence), Inclusion of Other in the Self Scale (Identification with their Ethnic Group and Identification with Canadians), Vancouver Index of Acculturation, and Mainstream vs Heritage Acculturation (Integration into Culture). Two corruption scores were calculated for each person using the mean perception of corruption index from Transparency International for all of the countries the participant have lived in and all countries from which they derived their ethnic heritage. The corruption index begins at 0 (most corrupt) up to 100 (least corrupt). For each country, we subtracted this value from 100 (so that higher scores indicated higher corruption). Perception of corruption was chosen as the measure of corruption as it indicated the perceived norm for national corruption.

The heritage corruption score primarily represents the potential influence of vertically transmitted corruption norms (parent to child), whereas the exposure corruption score represents corruption norms which the participant may have acquire through non-parental cultural transmission or direct experience.

We asked the last 39 groups (194 participants) their preferences for the conditions of the game. These participants were asked these questions after taking all other measures so that they were no different to the preceding 17 groups (79 participants). We report these preferences, along with details of all measures in the SI.

Data Availability Data can be downloaded at https://doi.org/10.6084/m9.figshare.5004956

References

1 Fehr, E. & Williams, T. *Endogenous emergence of institutions to sustain cooperation* (2013).

2 O'Gorman, R., Henrich, J. & Van Vugt, M. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 323-329 (2009).

3 Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466**, 861-863 (2010).

4 Ambraseys, N. & Bilham, R. Corruption kills. *Nature* **469**, 153-155 (2011).

5 Gächter, S. & Schulz, J. F. Intrinsic honesty and the prevalence of rule violations across societies. *Nature* (2016).

6 Weisel, O. & Shalvi, S. The collaborative roots of corruption. *Proceedings of the National Academy of Sciences* **112**, 10651-10656 (2015).

7 Kaufmann, D. Myths and realities of governance and corruption. *Available at SSRN 829244* (2005).

8 Treisman, D. What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annu. Rev. Polit. Sci.* **10**, 211-244 (2007).

9 Treisman, D. The causes of corruption: a cross-national study. *Journal of public economics* **76**, 399-457 (2000).

10 Gneezy, U., Saccardo, S. & van Veldhuizen, R. Bribery: Greed versus reciprocity. (Social Science Research Center Berlin (WZB), 2016).
11 Fehr, E. & Gachter, S. Cooperation and Punishment in Public Goods Experiments. *American Economic Review* **90**, 980-994 (2000).
12 Gächter, S., Renner, E. & Sefton, M. The long-run benefits of punishment. *Science* **322**, 1510-1510 (2008).
13 Rockenbach, B. & Milinski, M. The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718-723 (2006).
14 Henrich, J. *et al.* Costly punishment across human societies. *Science* **312**, 1767-1770 (2006).
15 Henrich, J. *et al.* Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480-1484 (2010).
16 Herrmann, B., Thöni, C. & Gächter, S. Antisocial Punishment Across Societies. *Science* **319**, 1362-1367, doi:10.1126/science.1153808 (2008).
17 Boyd, R. & Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and sociobiology* **13**, 171-195 (1992).
18 Panchanathan, K. & Boyd, R. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499-502 (2004).
19 Nikiforakis, N. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* **92**, 91-112, doi:http://dx.doi.org/10.1016/j.jpubeco.2007.04.008 (2008).
20 Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature* **466**, 29-29 (2010).
21 Traulsen, A., Röhl, T. & Milinski, M. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences*, rspb20120937 (2012).
22 Kenya, T. I. *The Kenya urban bribery index*. (Transparency International--Kenya, 2001).
23 Transparency International. (2014).
24 Olken, B. A. Corruption and the costs of redistribution: Micro evidence from Indonesia. *Journal of public economics* **90**, 853-870 (2006).
25 Pande, R. & Olken, B. Corruption in Developing Countries. *Annual Review of Economics* **4**, 479-509 (2012).
26 Chudek, M. & Henrich, J. Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences* **15**, 218-226 (2011).
27 Chudek, M., Muthukrishna, M. & Henrich, J. in *The Handbook of Evolutionary Psychology* Vol. 2 (ed David M. Buss) Ch. 30, (John Wiley and Sons, 2015).
28 Kimbrough, E. O. & Vostroknutov, A. Norms Make Preferences Social. (Department of Economics, Simon Fraser University, 2013).
29 Fisman, R. & Miguel, E. Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political economy* **115**, 1020-1048 (2007).
30 Kuan Yew, L. *From Third World to First World; the Singapore Story: 1965–2000*. (Harper Collins Publishers, 2000).

Acknowledgements J.H. acknowledges support from the Canadian Institute for Advanced Research. J.H. and P.F. acknowledge support from the Social Sciences and Humanities Research Council, Canada.

Author Contributions M.M., P.F., S.P., and J.H. developed theory, designed the experiments, and wrote the paper. M.M. and S.P. carried out the experiments. M.M., S.P., and J.H. conducted the statistical analyses. The authors declare no competing financial interests.