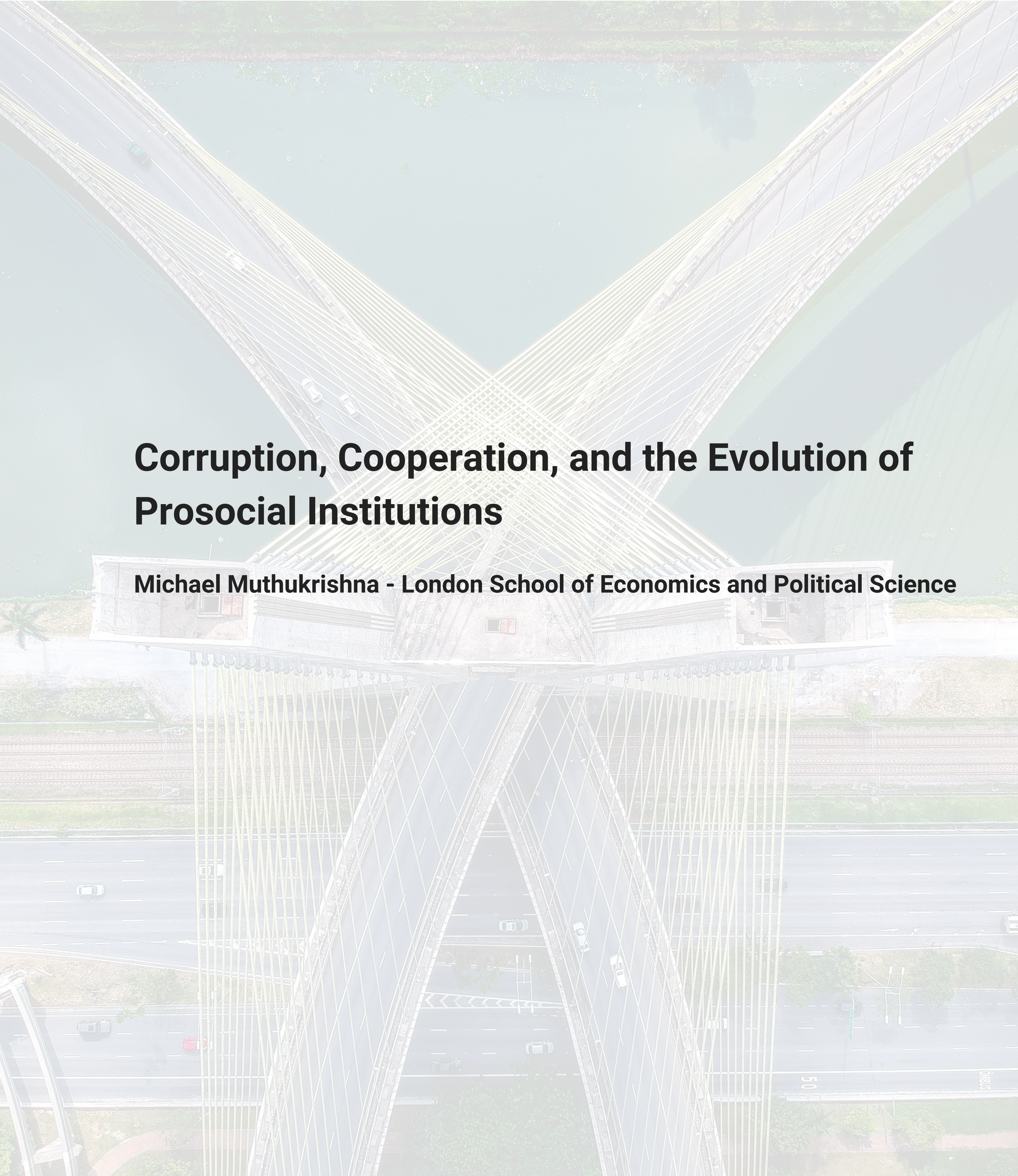
LSE Latin America and Caribbean Centre

Stanford University

Universidad de los Andes



Corruption, Cooperation, and the Evolution of Prosocial Institutions

Michael Muthukrishna - London School of Economics and Political Science

LSE-Stanford-Universidad de los Andes Conference on Long-Run Development in Latin America, London School of Economics and Political Science, 16-17 May 2018

Corruption, Cooperation, and the Evolution of Prosocial Institutions

Michael Muthukrishna*

May 2, 2018

Abstract

This article describes how corruption can and ought to be viewed as competing scales of cooperation. Viewing corruption through the lens of the cooperation literature gives us a mature theoretical and empirical framework from which to derive predictions and make sense of existing findings. This article was originally posted at Evonomics and ProMarket, The blog of the Stigler Center at the University of Chicago Booth School of Business as *Bribery, Cooperation, and the Evolution of Prosocial Instituions*. I have yet to expand the review beyond the 8 pages below, but I have attached a related empirical paper (Muthukrishna, et al., 2017, *Corruption cooperation and how anti-corruption strategies may backfire*) for discussion at the LSE-Stanford-Uniandes Conference.

*London School of Economics and Political Science

1 Corruption and Cooperation

There is nothing natural[1] about democracy. There is nothing natural about living in communities with complete strangers. There is nothing natural about large-scale anonymous cooperation. Yet, this morning, I bought a coffee from Starbucks with no fear of being poisoned or cheated. I caught a train on London's underground packed with people I've never met before and will probably never meet again. If we were commuting chimps in a space that small, it would have been a scene out of the latest Planet of the Apes by the time we reached Holborn station. We'll return to this mystery in a moment.

There is something very natural about prioritizing your family over other people. There is something very natural about helping your friends and others in your social circle. And there is something very natural about returning favors given to you. These are all smaller scales of cooperation that we share with other animals and that are well described by the math of evolutionary biology. The trouble is that these smaller scales of cooperation can undermine the larger-scale cooperation of modern states. Although corruption is often thought of as a falling from grace, a challenge to the normal functioning state—it's in the etymology of the word—it's perhaps better understood as the flip side of cooperation. One scale of cooperation, typically the one that's smaller and easier to sustain, undermines another.

When a leader gives his daughter a government contract, it's nepotism. But it's also cooperation at the level of the family, well explained by inclusive fitness[2], undermining cooperation at the level of the state. When a manager gives her friend a job, it's cronyism. But it's also cooperation at the level of friends, well explained by reciprocal altruism[3], undermining the meritocracy. Bribery is a cooperative act between two people, and so on. It's no surprise that family-oriented cultures like India and China are also high on corruption, particularly nepotism. Even in the Western world, it's no surprise that Australia, a country

[1]Putting aside what it means for something to be natural for our species, suffice to say these are recent inventions in our evolutionary history, by no means culturally universal, and not shared by our closest cousins.

[2]Genes that identify and favor copies of themselves will spread.

[3]Helping those who help you.

of mates, might be susceptible to cronyism. Or that breaking down kin networks predicts lower corruption and more successful democracies (Akbari and Kimbrough, 2017; Schulz, 2017). Part of the problem is that these smaller scales of cooperation are easier to sustain and explain than the kind of large-scale anonymous cooperation that we in the Western world have grown accustomed to.

So how is it that some states prevent these smaller scales of cooperation from undermining large-scale anonymous cooperation? The typical answer is that more successful nations have better institutions. All that's required is the right set of rules to make society function. But even on the face of it, this answer seems incomplete. If it were true, Liberia, who borrowed more than its flag from the United States, ought to be much more successful than it is[4]. Instead, these institutions are supported by invisible cultural pillars without which the institutions would fail. For example, without a belief in rule of law—that the law applies to all and cannot be changed on the whim of the leader—it doesn't matter what the constitution or legal code says, no one is listening. Without a long time horizon, decisions are judged on how well they serve our immediate needs making larger-scale projects, like reducing the effects of Climate Change, harder to justify[5]. Similarly, institutions often lack the punitive power to actually punish perpetrators. For example, most people in the US and UK pay their taxes, even though in reality the IRS and Her Majesty's Revenue and Customs lack the power to prosecute widespread non-compliance; your probability of getting caught is low. The tax compliant majority may never discover that they can cheat or how to get away with it (Chetty et al., 2013) and they may not actively seek this information as long as the probability of getting caught is non-zero, the system seems fair, and it seems like everyone else is complying. Or in other words, it's a combination of norms and institutions. But, it gets tricky—institutions are themselves hardened or codified norms[6] and the norms

[4]The United Nations Human Development Index ranks the United States 10th in the world. Liberia is 177th.

[5]Temporal discounting is the degree to which we value the future less than the present. Our tendency to value the present over the future is one reason we don't yet have moon or Mars colonies, but the degree to which we do this varies from society to society.

[6]Written laws can serve a signaling and coordination function; rather than having to interpret norms

themselves evolve in response to the present environment and due to path-dependence of previous environments, past decisions, and the places migrants come from. Modern groups vary on individualism (Talhelm et al., 2014) and even sexist attitudes (Alesina et al., 2013) based on their ancestors' farming practices[7].

The science of cultural evolution describes the evolution of these norms and introduces the possibility of out-of-equilibria behavior (people behaving in ways that do not benefit them individually) for long enough for institutions to try to stabilize the new equilibria. (For a summary of cultural evolution, see Henrich (2016) and for an even shorter summary see Chudek et al. (2015)).

2 How do we begin to understand these processes?

The real world is messy and before we start running randomized control trials or preparing case studies, it's useful to model the basic dynamics of cooperation using a simpler form that gets at the core elements of the challenge. One commonly used model is called the "Public Goods Game". The gist of the game is that I give you, and say 9 others, \$10. Whatever you put into a pool (the public good), I'll multiply by say 3, but then I'll divide the money equally regardless of contribution. This is similar to paying your taxes for public goods that we all benefit from, like roads, clean water, or environmental protections. The dilemma is this: the best move is for everyone to put all their money in the pool. Then they'll all go home with \$30. But it's in my best interests to put nothing in the pool and let everyone else put their money in. If I put in nothing and they put in \$10 each, I'll go home with almost \$40 ($\$10 \times 9 \times 3 \text{people}/10 = \$37$). What happens when we play this game?

Well, if we play it in a WEIRD[8] (Henrich et al., 2010) nation, where prosocial norms tend to be higher, people put about half their money in, but as they gradually realize

from the environment. When previously contentious norms are sufficiently well established, you may do well to codify them in law (legislating before they are established might mean more punishment—consider the history of prohibition in the United States).

[7]Not that agriculture is the main reason for these cultural differences!

[8]Western Educated Industrialized Rich Democratic

they can make more by putting in less, contributions dwindle to zero. One way to sustain contributions is to introduce peer punishment—allow people to spend some portion of their money to punish other people. This is similar to the kind of punishment we might see in a small village. I know who you are or at least I know your parents or people you know. If you steal my crops, I'll punish you myself or ruin your reputation. In the game, if we introduce the possibility of peer punishment, contributions rise again. The problem is that this doesn't scale well. As the number of people grows, we get second-order free-riding—people prefer to let someone else pay the cost of punishment. When someone cuts a queue, you grumble—someone ought to tell that person off! Someone other than me... And you can also get counter-punishment—revenge for being punished. The best solution seems to be to create a punishment institution. Pick one person as a "Leader" and allow them to extract taxes that can be used to punish free-riders. This works really well and scales up nicely. It's similar to a functioning police force and judiciary in WEIRD nations. In fact, the models suggest that the more power you give to the leader, the more cooperation they can sustain. Aha! Problem solved. Not quite. Models like these are very useful for distilling the core of a phenomenon, they can miss things. Recall where we started—smaller-scales of cooperation can undermine the larger-scale.

In Muthukrishna, Francois, et al. (2017), we wanted to show just how easy it was to break that well-functioning institution. We did it by introducing the possibility of another very simple form of cooperation—you scratch my back, I'll scratch yours—bribery. And then we wanted to show the power of invisible cultural pillars by measuring people's cultural background and by trying to fix corruption using common anti-corruption strategies. We wanted to show that these strategies, including transparency, don't work in all contexts and can even backfire.

Our *Bribery Game* was the usual institutional punishment public goods game with the punishing leader, but with one additional choice—players could not only keep money for themselves or contribute to the public pool, they could also contribute to the leader. And the leader could not only punish or not punish, they could instead accept that contribution. What happened? On average, we saw contributions fall by 25% compared to the game without bribery as an option. More than double what the pound has fallen against the USD since Brexit (∼12%). Fine, bribery is costly. The World Bank estimates $1 trillion is paid in bribes alone; in Kenya, 8 out of 10 interactions with public officials involves a bribe, and as Manfred Milinski (2017) points out in his summary of our paper, most of humanity—6 billion people—live in nations with high levels of corruption. Our model also reveals that unlike the typical institutional punishment public goods game, where stronger institutions mean that more cooperation can be sustained, when bribery is an option, stronger institutions mean more bribery. A small bribe multiplied by the number of players will make you a lot richer than your share of the public good!

3 So can we fix it?

The usual answer is transparency. There are also some interesting approaches, like tying a leader's salary to the country's GDP—the Singaporean model[9]. So what happened when we introduced these strategies? Well, when the public goods multiplier was high (economic potential—potential to make money using legitimate means—was high) or the institution had power to punish, then contributions went up. Not to levels without bribery as an option, but higher. But in poor contexts with weak punishing institutions, transparency had no effect or backfired. As did the Singaporean model[10]. Why?

Consider what transparency does. It tells us what people are doing. But as psychological and cultural evolutionary research reveals, this solves a common knowledge problem and reveals the descriptive norm—what people are doing. For it to have any hope of changing behavior, we need a prescriptive or proscriptive norm against corruption. Without this,

[9] Singapore's leaders are the highest paid in the world, but the nation also has one of the lowest corruption rates in the world—lower than the Netherlands, Canada, Germany, UK, Australia, and United States [source: https://www.transparency.org/news/feature/corruption_perceptions_index_2016].

[10] Note, there are some conceptual issues that make interpretation of the Singaporean treatment ambiguous. We discuss this in the supplementary. We'll have to further explore this in a future study.

transparency just reinforces that everyone is accepting bribes and you'd be a fool not to. People who have lived in corrupt countries will have felt this frustration first hand. There's a sense that it's not about bad apples—the society is broken in ways that are sometimes difficult to articulate. But societal norms are not arbitrary. They are adapted to the local environment and influenced by historical contexts. In our experiment, the parameters created the environment. If there really is no easy way to legitimately make money and the state doesn't have the power to punish free-riders, then bribery really is the right option. So even among Canadians, admittedly some of the nicest people in the world, in these in-game parameters, corruption was difficult to eradicate. When the country is poor and the state has no power, transparency doesn't tell you not to pay a bribe, it solves a different problem—it tells you the price of the bribe. Not "should I pay", but "how much"?

There were some other nuances to the experiment that deserve follow up. If we had played the game in Cameroon instead of Canada, we suspect baseline bribery would have been higher. Indeed, people with direct exposure to corruption norms encouraged more corruption in the game controlling for ethnic background. And those with an ethnic background that included more corrupt countries, but without direct exposure were actually better cooperators than the 3rd generation+ Canadians. These results may reveal some of the effects of migration and historical path dependence. Of course, great caution is required in applying these results to the messiness of the real world. We hope to further investigate these cultural patterns in future work.

The experiment also reveals that corruption may be quite high in developed countries, but its costs aren't as easily felt. Leaders in richer nations like the United States may accept "bribes" in the form of lobbying or campaign funding and these may indeed be costly for the efficiency of the economy, but it may be the difference between a city building 25 or 20 schools. In a poor country similar corruption may be the difference between a city building 3 or 1 school. Five is more than 3, but 3 is three times more than 1. In a rich nation, the cost of corruption may be larger in absolute value, but in a poorer nation, it may be larger in relative value and felt more acutely.

The take home is that cooperation and corruption are two sides of the same coin; different scales of cooperation competing. This approach gives us a powerful theoretical and empirical toolkit for developing a framework for understanding corruption, why some states succeed and others fail, why some oscillate, and the triggers that may lead to failed states succeeding and successful states failing.

Our cultural evolutionary biases lead us to look for whom to learn from and perhaps whom to avoid. They lead us to blame individuals for corruption. But just as atrocities are the acts of many humans cooperating toward an evil end, corruption is a feature of a society not individuals.

Indeed, corruption is arguably easier to understand than my fearless acceptance of my anonymous barista's coffee. Our tendency to favor those who share copies of our genes—a tendency all animals share—lead to both love of family and nepotism. Putting our buddies before others is as ancient as our species, but it creates inefficiencies in a meritocracy. Innovations are often the result of applying well-established approaches in one area to the problems of another (Muthukrishna and Henrich, 2016). We hope the science of cooperation and cultural evolution will give us new tools in combating corruption.

References

Akbari, M. & Kimbrough, E. O. (2017). Kinship, Fractionalization and Corruption. doi:10.2139/ssrn.2847222

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the Origins of Gender Roles: Womenand the Plough. *The Quarterly Journal of Economics*, *128*(May), 469–530. doi:10.1093/qje/qjt005.Advance

Chetty, R., Friedman, J. N., & Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review*, *103*(7), 2683–2721. doi:10.1257/aer.103.7.2683

Chudek, M., Muthukrishna, M., & Henrich, J. (2015). Cultural evolution. In *The handbook of evolutionary psychology*. Wiley Online Library.

Henrich, J. (2016). The secret of our success: How learning from others drove human evolution, domesticated our species, and made us smart. Princeton University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and brain sciences*, *33*(2-3), 61–83. doi:10.1017/S0140525X0999152X

Milinski, M. (2017). Economics: Corruption made visible. *Nature Human Behaviour*, *1*(7), 0144. doi:10.1038/s41562-017-0144

Muthukrishna, M., Francois, P., Pourahmadi, S., & Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behaviour*, *1*(7), 0138. doi:10.1038/s41562-017-0138

Muthukrishna, M. & Henrich, J. (2016). Innovation in the collective brain. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *371*(1690), 137–148. doi:10.1098/rstb.2015.0192

Schulz, J. F. (2017). The Churches' Bans on Consanguineous Marriages, Kin-Networks and Democracy. *SSRN Electronic Journal*, (2877828), 1–58. doi:10.2139/ssrn.2877828

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science (New York, N.Y.) 344*(6184), 603–8. doi:10.1126/science.1246850

nature human behaviour

LETTERS

PUBLISHED: 10 JULY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0138

Corrupting cooperation and how anti-corruption strategies may backfire

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

Understanding how humans sustain cooperation in large, anonymous societies remains a central question of both theoretical and practical importance. In the laboratory, experimental behavioural 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 (a 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 (1) corruption possibilities cause a large (25%) decrease in public good provisioning, (2) empowering leaders decreases cooperative contributions (in direct opposition to typical institutional punishment results), (3) growing up in a more corrupt society predicts more acceptance of bribes and (4) anti-corruption strategies are effective under some conditions, but can further decrease public good provisioning when leaders are weak and the 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 (although not in small-scale societies), there is some variability in both the motivation to punish free-riders and the tendency to punish cooperators (for instance, some societies display significant levels of antisocial punishment—the punishment of cooperators)[14–16].

Research on the role of peer punishment in sustaining cooperation reveals two major challenges: (1) the second-order free-rider problem in which individuals defect on the job of punishing and thereby increase their payoffs[17,18] and (2) 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 participants from western, educated, industrialized, rich and democratic (WEIRD) nations[20]) prefer institutional punishment to peer punishment[1,21].

Institutional punishment, as typically modelled in public goods games (PGGs), serves to incentivize player choices when contributing to the public pool, and works 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 (for example, by evading taxes) and to avoid being punished (for example, by paying a bribe instead). In other words, real-world 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, the 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 per month[22]. In contrast, the average Dane may never pay a bribe in their lifetime as 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 programmes[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, with recent estimates placing its levels of corruption close to those of China and Brazil[23]. Corruption in European Union states, such as Greece, potentially undermines the future of the European Union. Although 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). In a PGG, 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 the players regardless of contribution. A cooperative dilemma is created by setting the multiplier such that it is in every player's best interest to allow others to contribute while 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

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science, London WC2A 2AE, UK. [2]Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Vancouver School of Economics, University of British Columbia, Vancouver, British Columbia V6T 1L4, Canada. [4]Canadian Institute for Advanced Research, Toronto, Ontario M5G 1M1, Canada. [5]Department of Economics, Columbia University, New York, New York 10027, USA. [6]Department of Psychology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. *e-mail: m.muthukrishna@lse.ac.uk

Table 1 | Leader decisions based on economic potential, leader strength and corruption exposure scores.

	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)
Observations	1,396	1,396	1,396
n	175	175	175
Groups	45	45	45
Deviance information criterion	36.13	18.23	18.45

Values are reported as odds ratios and highest posterior density 95% confidence intervals. Odds ratios were estimated using a Markov chain Monte Carlo categorical generalized linear mixed model regression with the behaviour coded as 1 and the other two behaviours coded as 0. Each model regressed the behaviour in the BG (with no transparency or leader investment) on economic potential (low versus high), leadership strength (weak versus 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 the Supplementary Information.

as a leader who can allocate punishments using taxes extracted from other players. Past research has shown the effectiveness of assigning 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). In this scenario, in addition to dividing their endowment between themselves and the public pool, players can also offer some of their endowment to improve the leader's payoff (that is, effectively offering a bribe, although we use neutral language). In turn, leaders have an additional exclusive choice in addition to punishing or doing nothing to players: they can choose to take the contribution (that is, 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 would not last long. A new leader was selected in each round to remove any reputational effects, which turned the game into a series of repeated one-shot encounters. We manipulated the pool multiplier (a proxy for economic potential) and the punishment multiplier (the power of the leader to punish). In the BG, we also introduced three corruption mitigation strategies: partial transparency (revealing leader contributions), full transparency (revealing all leader behaviour, 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, which included native-born Canadians and first- and second-generation immigrants with diverse backgrounds.

We assumed players: (1) brought cultural differences to the game, which were shaped by their different ethnic backgrounds and cultural exposure; and (2) adjusted their behaviours via exposure to the experimental setting, moving closer to the equilibrium that maximized payoffs. We modelled an IPGG with a fixed tax rate to more realistically capture a world in which taxes were not directly correlated with punishment and where leaders could punish without a large cost to themselves (since their own taxes were a small part of the taxes contributing to the pool punishment or institution). We then modified the game to turn it into a BG by offering players and leaders the choice to offer and accept bribes. Without punishment, contributions tend towards zero. This is because contribution levels are contingent on the strength of leaders and their tendency to punish low contributors. We predicted that leaders would use taxes as punishment in the IPGG, since they are not personally costly and they benefit the leader's payoff by increasing the size of the public good. With increased leader strength, we predicted higher contributions and more public good provisioning. With regards to the BG, we predicted that players would have no incentive to offer contributions or bribes unless they were punished for not doing so. However, when bribery was an option, leaders would have a greater incentive to punish people for not offering brides than for not contributing, since their share of the public good would 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 predicted that stronger leaders in the BG would reduce contributions and public good provisioning. However, if players had a preference for contributions over bribes (for example, if their previous experience was a world where potential returns on the public good were higher or where anti-corruption norms were adaptive), the incentive to punish bribes over contributions would be dampened. In contrast, growing up in a more corrupt society may lead to a higher preference for eliciting, offering and accepting bribes. Our full set of predictions is provided in the Supplementary Information.

To examine the costs of corruption, we compared the IPGG and BG. We found that when bribery was an option, mean contributions dropped by 25%. The difference between these conditions (estimated using a Markov chain Monte Carlo generalized linear mixed model regression; Supplementary Table 2) represented a 0.43 (95% confidence interval: −0.49 to −0.38) s.d. loss (1.4 points per period, equivalent to 14% of the initial endowment or Canadian $2.10 over the course of the game). Not surprisingly, when corruption could enter, it did, and cooperation deteriorated.

Having established the impact of bribery on cooperation, we examined the causes of this corruption. In Table 1 and Fig. 1 we used a Markov chain Monte Carlo categorical generalized linear mixed model regression 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 (that is, stronger leaders) were about twice as likely to accept bribes and about three times less likely to do nothing. In contrast, when accepting bribes was not an option (that is, in the IPGG), the more powerful leaders were as likely to do nothing (see 'Leader decisions' in Supplementary Information). Thus, as expected, more power led to more corrupt behaviour.

Exploring individual variation, we found that those who grew up in more corrupt countries were more willing to accept bribes. For every one s.d. increase in players' exposure corruption scores (see 'Corruption perception scores' in Supplementary Information for

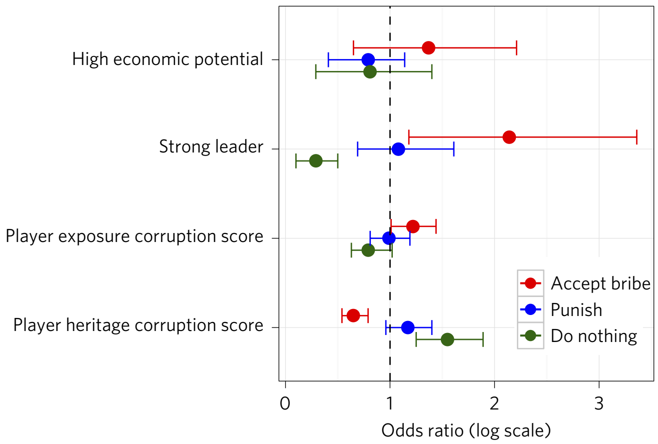


Figure 1 | Leader decisions based on economic potential, leader strength and corruption exposure scores. Odds ratios and 95% confidence intervals are shown for each behaviour (accept bribe, punish or do nothing).

		Weak leaders: Control	Weak leaders: BG	Strong leaders: Control	Strong leaders: BG
Poor economic potential	Control		0.21***		0.52****
	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****
Rich economic potential	Control		0.39****		0.57****
	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****

Figure 2 | Cures for corruption when there is a weak versus strong leader and when there is rich versus poor economic potential. Darker blue indicates greater public goods provisioning and darker red indicates reduced public goods provisioning. All coefficients were extracted from a single model by changing reference groups. The columns represent the reference group treatment (control versus BG), while each row shows the coefficient of each treatment compared with this reference group. The contributions were *z* scores, so the coefficients represent s.d. The full model is reported in the Supplementary Information. In all models, we accounted 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 (although not to control levels) when leaders are strong or the economic potential is rich. When leaders are weak and the economic potential is poor, the apparent corruption mitigation strategy, full transparency has no effect and partial transparency further decreases contributions. $^{*}P < 0.10$; $^{**}P < 0.05$; $^{***}P < 0.01$; $^{****}P < 0.001$.

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 (that is, more perceived corruption), leaders were 1.5 times less likely to accept bribes for every s.d. increase in corruption score and 1.6 times more likely to do nothing (see Fig. 1; the Supplementary Information shows all the models). In combination with other evidence[5,6,26–29], we suspect that our corruption exposure scores captured internalized social norms related to corruption acquired while growing up in different communities. Meanwhile, our parental heritage effects, which were driven by the Canadian-born participants (for example, second-generation immigrants), may have captured an internalized reaction against ethnic stereotyping—for instance, a reaction against the assumption of corrupt behaviour from those of their heritage[21].

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 was explicitly used as one aspect of an anti-corruption measure in Singapore, which has one of the lowest levels of corruption (based on the Corruption Perceptions 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. The leader investment treatment was designed to be similar to linking leader payoffs to a country's gross domestic product, but in a way that minimally deviated 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 BG. We regressed contributions (*z* scores) on treatment, economic potential and leader strength. The results of this regression are shown in Fig. 2 and separate coefficients within each condition can be seen. Note that these values come from a single model and were calculated by changing reference groups (see Supplementary Information). The raw mean contribution values are shown in Fig. 3.

Figures 2 and 3 reveal that stronger leaders were better able to increase the efficiency of public goods provisioning when the economic potential was poor and bribery was not an option (red bars in the top row), but when bribery was an option (blue bars) stronger leaders in poor contexts reduced the efficiency of the public good, making themselves wealthy at the expense of other players. Corruption mitigation effectively increased contributions (although not to control levels) when leaders were strong or the economic potential was rich. When leaders were weak and the economic potential was poor, the apparent corruption mitigation strategy, full transparency, had no effect and partial transparency further decreased contributions to levels lower than the standard BG (leading 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 were accepting bribes at comparable levels in both poor and rich economic contexts, the degree of corruption was not as visible if the economic potential was high. Leaders in richer economic contexts, such as 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 is not as obvious since the economic potential is already much higher than in other nations. In contrast, in poorer economic contexts, such as the Democratic Republic of the Congo,

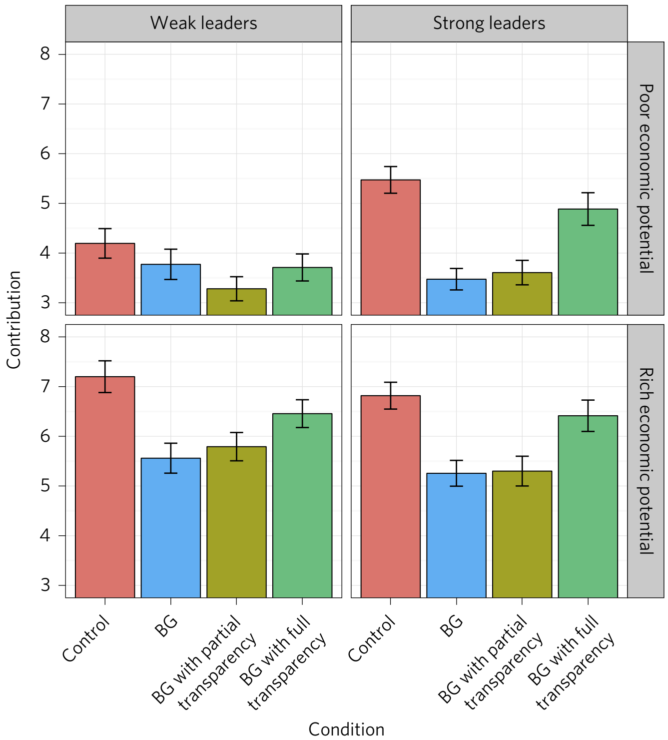


Figure 3 | Leader contributions by condition. Raw contributions (of the ten endowed points) and 95% confidence intervals for each within-subject treatment (control, BG, BG with partial transparency or BG with full transparency) in each between-subjects structural context (strong versus weak leader and poor versus rich economic potential). These data are consistent with our theory that predicts that more powerful leaders increase contributions in the IPGG but decrease contributions in the BG.

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 could worsen the situation.

Our results suggest that the effect of exposure to different institutions and norms persists after moving to a new environment. This increase in corrupt behaviour following 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 used a large range of corruption scores (see 'Corruption perception scores in Supplementary Information), our sample was limited to migrants in a Canadian context and further investigation is required to determine if these cultural results can be generalized. Together, these results suggest that corruption may be rooted in structural factors, but that internalized corruption norms may cause these behaviours to persist in a new context.

Overall, these results suggest that: (1) stronger institutions and leaders are required to sustain public goods contributions when the economic potential is poor and the incentive to free ride is high; (2) in this context, when they are able to, leaders abuse their power with a noticeable economic cost; and (3) despite this, even if the 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 as they reveal the rationality of low public good contributions and show 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 equilibrium: bribe offers and low public good provisioning.

Although 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 investigations, our results suggest that as the 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 the 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. A total of 274 participants (166 females; mean age: 20.90), drawn from an economic subject pool open to the public, took part in the study. Their ethnic backgrounds were as follows: 63 European Canadians, 158 East Asians, 17 South Asians and 36 of other ethnicities. The participants played in groups of between four and seven players. Ethical approval was obtained from the University of British Columbia Behavioural Research Ethics Board (H12-02457). Informed consent was obtained from all participants before the start of the study. The participants were randomly assigned to the experimental groups.

Experimental design. We used a 2 (high versus low economic potential) × 2 (weak versus strong leader power) between-subjects experimental design with five within-subject treatments: IPGG control ($n = 205$), BG ($n = 222$), BG with partial transparency ($n = 228$), BG with full transparency ($n = 204$) and BG with leader investment ($n = 196$). Allocation to all treatments was random. The sample sizes for the four between-subjects treatments were as follows: low economic potential and weak leader power ($n = 71$), low economic potential and strong leader power ($n = 68$), high economic potential and weak leader power ($n = 68$) and high economic potential and strong leader power ($n = 67$).

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 four between-subjects treatments and four of the five within-subjects treatments.

To test the possible contributing causes of corruption, we randomly assigned each group of participants to a treatment with (1) either a high or low marginal per capita rate of return (0.3 versus 0.6) as a measure of economic potential and (2) either a high or low punishment multiplier (1 versus 3) as a measure of the strength of the leader or institution. The marginal per capita rate of return was the expected return for every point invested in the public pool and the punishment multiplier was the number of points subtracted from a sanctioned player for every tax point spent on punishing that player.

The within-subject treatments were played in a random order with pre-recorded video instructions before each period. A quiz was conducted at the start to ensure participants knew how each treatment worked. This quiz, along with the script and screenshots from the video, is in the Supplementary Information.

We used a block randomization design, in which participants played a minimum of ten rounds, but the game could end at any point before the completion of ten rounds. At ten rounds, the participants were informed which round the period had ended at or played further rounds until the game ended. In this way, there were ten rounds to analyse without end-game effects—that is, participants did not know when the game would end. To remove reputational effects, the leader was also randomly selected for each round. Replacement was performed by random selection, such that players also could not say that the same person could not be the leader for a consecutive round. As such, the experiment could be interpreted as a series of one-shot interactions. The participants were paid for ten random rounds from across all the 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 participants had spent their entire life in Canada, where else they had lived, which suburb they had grown up in, ethnic group, religion and importance of religion, how well they spoke the language of their ethnic heritage (cultural competence), inclusion of other in the self scale (identification with their ethnic group and identification with Canadians), the Vancouver Index of Acculturation, and mainstream versus heritage acculturation (integration into culture). Two corruption scores were calculated for each participant using the mean of Transparency International's Corruption Perceptions Index for all the countries each participant had lived in and all the countries from which they derived their ethnic heritage. The Corruption Perceptions Index has a scale from 0 (most corrupt) 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 that the participant may have acquired 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 all other measures had been taken so that there were no differences in experimental design between them and the preceding 17 groups (79 participants). We report these preferences, along with the details of all the measures in the Supplementary Information.

Data availability. The data that support the findings of this study are available in figshare with the identifier https://doi.org/10.6084/m9.figshare.5004956.

Received 13 July 2016; accepted 7 June 2017; published 10 July 2017

References

1. Fehr, E. & Williams, T. *Endogenous Emergence of Institutions to Sustain Cooperation* (2013); http://researchers-sbe.unimaas.nl/wp-content/uploads/gsbe/spring-2014/papers-and-abstracts/Paper_williams.pdf
2. O'Gorman, R., Henrich, J. & Van Vugt, M. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proc. R. Soc. Lond. B Biol. Sci.* **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* **531,** 496–499 (2016).
6. Weisel, O. & Shalvi, S. The collaborative roots of corruption. *Proc. Natl Acad. Sci. USA* **112,** 10651–10656 (2015).
7. Kaufmann, D. Myths and realities of governance and corruption. http://dx.doi.org/10.2139/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. *J. Pub. Econ.* **76,** 399–457 (2000).
10. Gneezy, U., Saccardo, S. & van Veldhuizen, R. Bribery: greed versus reciprocity. http://dx.doi.org/10.2139/ssrn.2803623 (2016).
11. Fehr, E. & Gachter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **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 (2008).
17. Boyd, R. & Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **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? *J. Pub. Econ.* **92,** 91–112 (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. *Proc. R. Soc. B Biol. Sci.* **279,** 3716–3721 (2012).
22. Kenya, T. I. *The Kenya Urban Bribery Index* (Transparency International—Kenya, 2001).
23. Transparency International. Corruption perceptions index 2014 brochure. https://www.transparency.org/cpi2014/results#myAnchor2 (2014).
24. Olken, B. A. Corruption and the costs of redistribution: micro evidence from Indonesia. *J. Pub. Econ.* **90,** 853–870 (2006).
25. Pande, R. & Olken, B. Corruption in developing countries. *Annu. Rev. Econom.* **4,** 479–509 (2012).
26. Chudek, M. & Henrich, J. Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cogn. Sci.* **15,** 218–226 (2011).
27. Chudek, M., Muthukrishna, M. & Henrich, J. in *The Handbook of Evolutionary Psychology* Vol. 2 (ed. Buss, D. M.) Ch. 30 (John Wiley and Sons, 2015).
28. Kimbrough, E. O. & Vostroknutov, A. *Norms Make Preferences Social* (Dept of Economics, Simon Fraser Univ., 2013).
29. Fisman, R. & Miguel, E. Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets. *J. Polit. Econ.* **115,** 1020–1048 (2007).
30. Kuan Yew, L. *From Third World to First World; the Singapore Story: 1965–2000* (Harper Collins, 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. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

M.M., P.F., S.P. and J.H. developed the 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.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M.

How to cite this article: Muthukrishna, M., Francois, P., Pourahmadi, S. & Henrich, J. Corrupting cooperation and how anti-corruption strategies may backfire. *Nat. Hum. Behav.* **1,** 0138 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Competing interests

The authors declare no competing interests.

Corrupting Cooperation and How Anti-Corruption Strategies May Backfire
Supplementary Information

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

Table of Contents

Supplementary Methods

Experimental Design

Random Assignment to Treatments

Each group of players was randomly assigned to:

1. Version A or B, which determined whether demographics and background questions were administered before or after game play.
2. Low or High economic potential, which was the public goods game marginal per capita return (MPCR).
    - High Economic Potential: 0.6 MPCR

    - Low Economic Potential: 0.3 MPCR
3. Weak or Strong leader power, which was the punishment multiplier for institutional punishment.
    - Weak Leader Power: 1 x punishment multiplier
    - Strong Leader Power: 3 x punishment multiplier
4. Four of the five versions of the public goods game (institutional punishment public goods game; control, bribery game, bribery game with partial transparency, bribery game with full transparency, bribery game with leader investment) to be played in a random order. Four rather than five treatments were administered due to time constraints. Economic potential and leader power did not change between games.

Random assignment was performed using the Random functions in Excel or Google Sheets. Since the experimenter selected the treatments in the software, they were not blind to the treatment.

In the partial transparency treatment, players could see the leaders contribution to the public pool. This offered leaders an opportunity to reveal a norm through their own actions.

In the full transparency treatment, players could see all leader actions (though anonymized)—the size of contributions, the size of bribes, and the leaders decision in each case. This level of transparency in leader behaviors would be the ultimate goal of many campaigns and can be explored in an experimental setting.

In the leader investment treatment, leaders are forced to maximally contribute to the public pool tying their own payoff to its success. This treatment was inspired by the Singaporean model. In the Singaporean model, a leaders' salary is determined by (a) the highest paid professionals in the country and (b) the GDP of the country. The goal here is to incentivize leaders to increase average and top-end growth in society. Although it could be that higher salaries rather than the direct incentives reduce corruption in Singapore, recent evidence from Ghana and neighboring countries suggests that increasing police officer salaries actually *increases* bribe behavior[1]. Ideally, to fully capture the Singaporean model, we would have included some kind of extra bonus based on the size of the public pool and perhaps the payoff to the highest earning player, but we didn't want this treatment to be too far out of line with the other treatments. That is, we wanted to test leader investment with minimal changes to the overall game structure. Nevertheless, in its current form, the effect of this treatment has alternative explanations. Most concerning, forcing leaders to maximally contribute, leads to them having large stake in the public goof, a particular problem in smaller groups. For this reason, without further investigation, we should be especially cautious about the degree to which these results will translate to larger populations and we only report the results in the Supplementary Information.

In order to remove reputational effects, for each round, the leader was also randomly selected by the software, with replacement, such that leaders could be leaders for consecutive rounds.

Procedure

Participants entered the room and were asked to sit down at a computer. Computers were separated by a barrier so that participants could not see other players' screens. A consent form, pen and headphones were laid out on each keyboard. Participants were told to put on the headphones and click play on the video. All further instructions were provided via the software and via video to ensure all participants received the same information. Sample screenshots for each treatment are shown in the next section (Screenshots from Experiment) and full scripts with screenshots from the video are shown in the final section (Experimental Protocol). After the instructions for each treatment, participants were administered a quiz via the experimental software. The quiz ensured that participants understood the instructions. Participants had to answer questions correctly to begin the treatment. The quiz questions are also provided in the final section (Experimental Protocol).

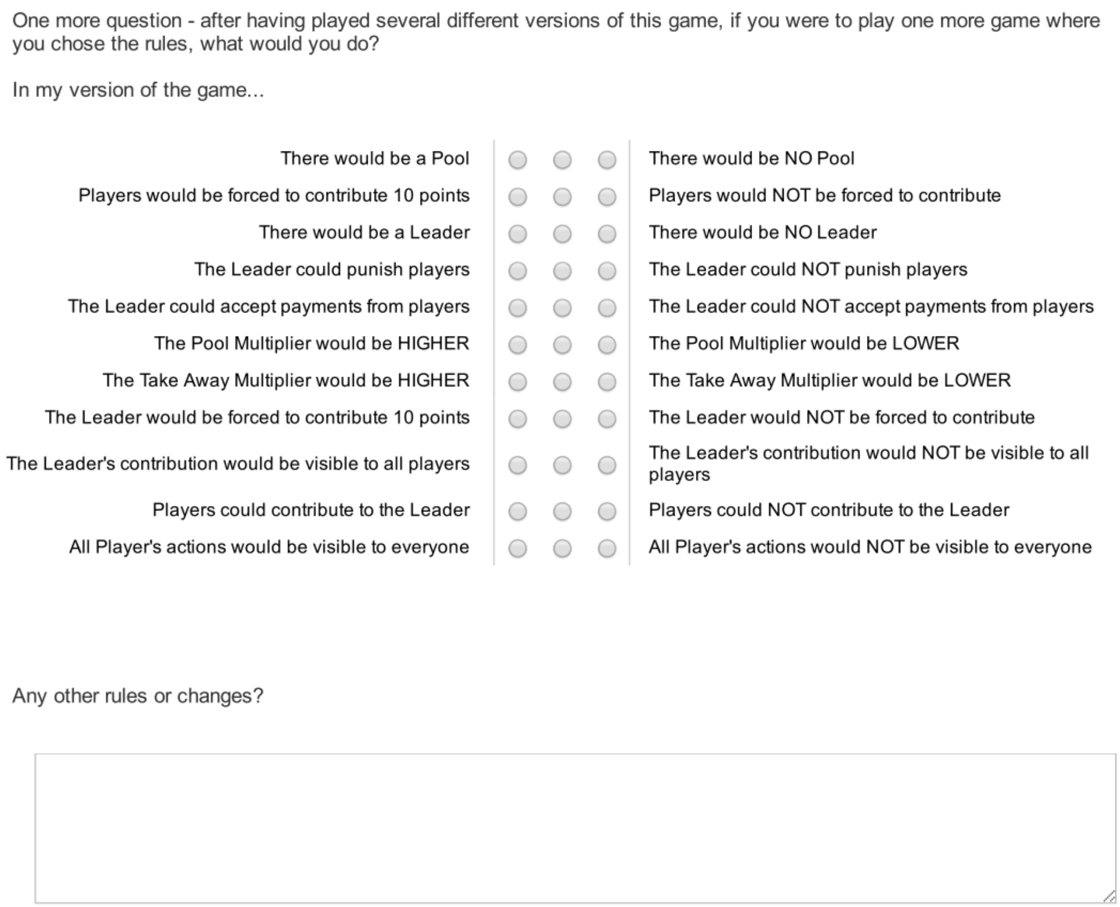
Measures Collected

In addition to player and leader behavior in the game, we collected the following measures. This is a complete list of all measures collected. No additional measures were collected.

1. Prestige and Dominance Scale [Self-report version][2]
2. Right Wing Authoritarianism (RWA) scale[3]
3. How old are you in age?
4. What is your gender?
5. If you are student, what degree are you studying for (e.g. B Arts, B Sc)? If you are working, what is your occupation (e.g. Pharmacist)?
6. Major (if degree) What is your major (e.g. Chemistry) or industry (e.g. Health)?
7. Have you lived your entire life in Canada?
8. If no, where else have you lived (please list)? [Note: these countries were used to calculate the Exposure Corruption Score]
9. What suburb do/did you live in for most of your time in Canada?
10. Please specify the ethnic (cultural) group you primarily identify with (e.g. Punjabi,Cantonese Chinese, Mandarin Chinese, Japanese, European, etc.) [Note: these identities were used to calculate the Heritage Corruption Score. Cantonese Chinese were assumed to be from Hong Kong and Mandarin Chinese were assumed to be from China. Ambiguous country of origin, such as Armenian, were not included]
11. What is the native language of your ethnic group?
12. How well do you speak the native language of your ethnic group?
13. Inclusion of Other in the Self Scale[4] for ethnic group

14. Inclusion of Other in the Self Scale[4] for other Canadians
15. What is your religious background?
16. How important is religion in your daily life?
17. Vancouver Index of Accultaration[5]

The last 39 groups (194 participants) were also asked the following questions about their preferences for the game:



One more question - after having played several different versions of this game, if you were to play one more game where you chose the rules, what would you do?

In my version of the game...

There would be a Pool	There would be NO Pool
Players would be forced to contribute 10 points	Players would NOT be forced to contribute
There would be a Leader	There would be NO Leader
The Leader could punish players	The Leader could NOT punish players
The Leader could accept payments from players	The Leader could NOT accept payments from players
The Pool Multiplier would be HIGHER	The Pool Multiplier would be LOWER
The Take Away Multiplier would be HIGHER	The Take Away Multiplier would be LOWER
The Leader would be forced to contribute 10 points	The Leader would NOT be forced to contribute
The Leader's contribution would be visible to all players	The Leader's contribution would NOT be visible to all players
Players could contribute to the Leader	Players could NOT contribute to the Leader
All Player's actions would be visible to everyone	All Player's actions would NOT be visible to everyone

Any other rules or changes?

Figure S1. After the experiment had concluded, participants were asked for their preferred game paramters.

Corruption Perception Scores

Below are histograms for distributions of heritage corruption score (the mean of the Corruption Perception Index[6] values of players' countries of ethnic heritage) and a exposure corruption score (the mean of the Corruption Perception Index values of the countries in which they had lived). As discussed in the main text, the heritage corruption score represents the potential influence of vertically transmitted corruption norms (parent to child), whereas the exposure corruption score represents corruption norms to which the participant was directly exposed (i.e., potentially personal experience as well as vertical, horizontal, and oblique transmission).

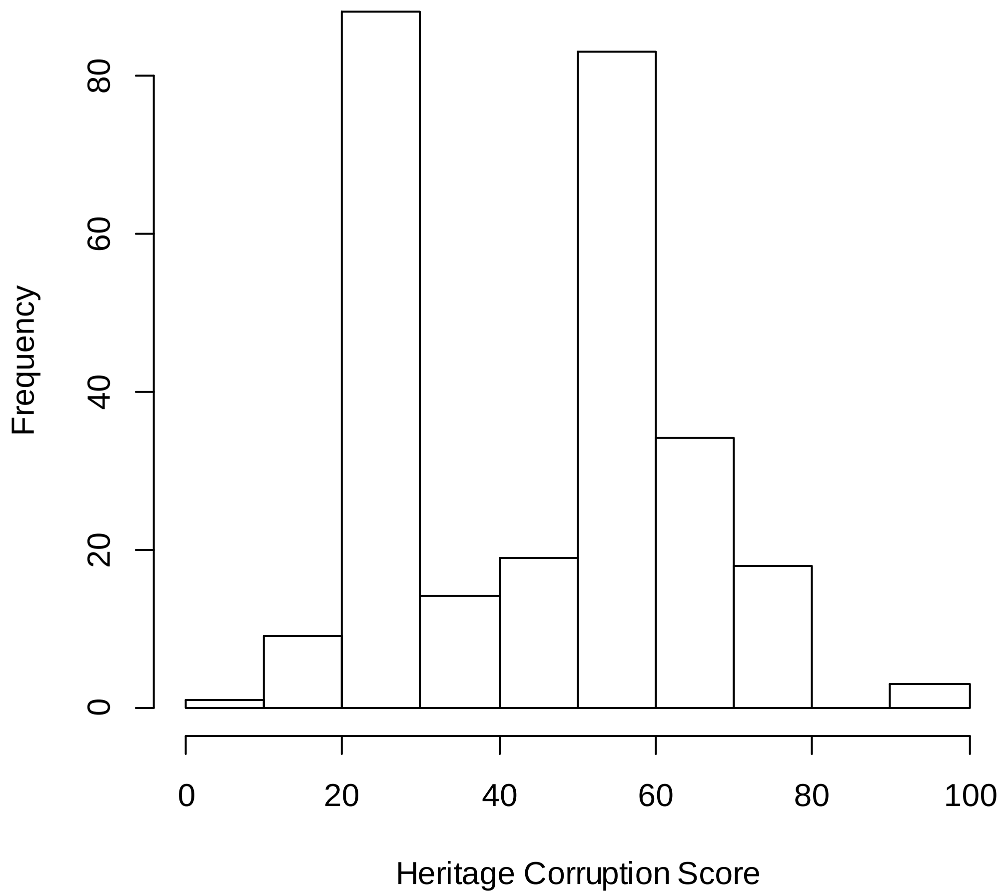


Figure S2. Histogram of Heritage Corruption Scores.

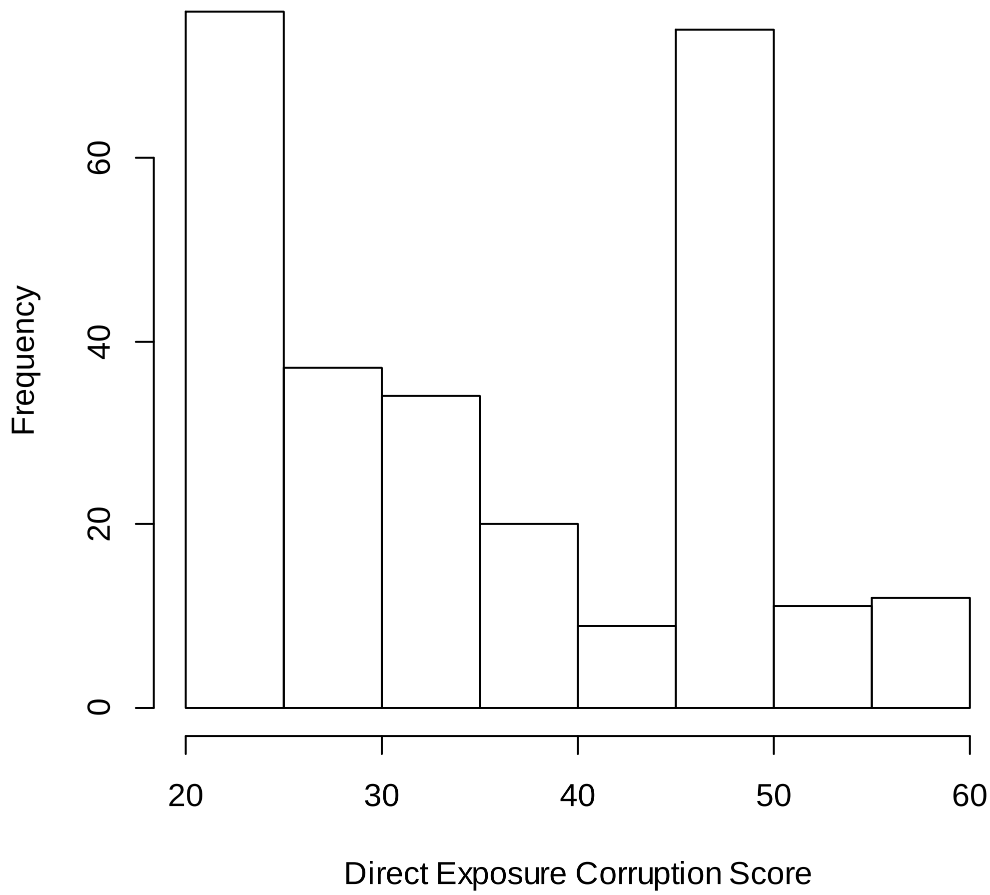


Figure S3. Histogram of Exposure Corruption Score.

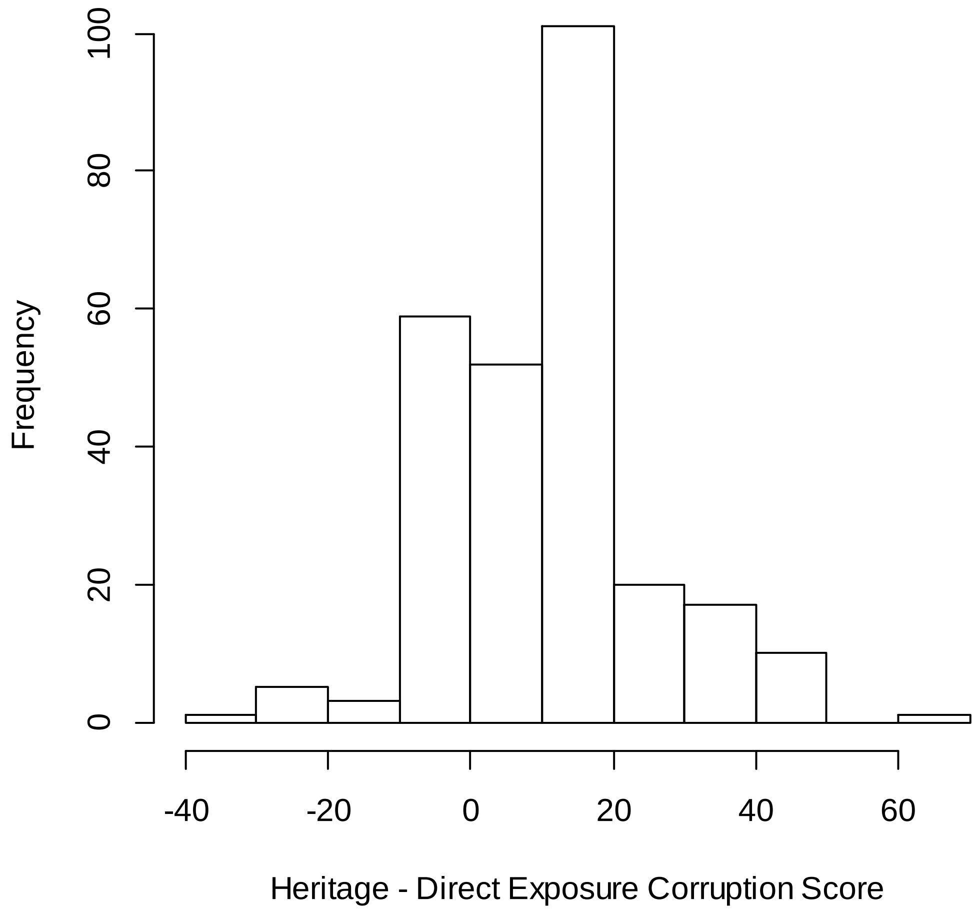


Figure S4. Histogram of Exposure Corruption Score subtracted from Heritage Corruption Score. This plot illustrates that these scores are not identical and in some cases, have very different values.

Sampling

Based on a pilot of 50 participants, we expected a fairly large effect size and aimed to have at least 250 participants. All data was included, unless the software crashed (resulting in data loss or less than 10 rounds for that treatment). This happened for 5 of the 224 treatment groups. Incomplete data (and obviously lost data) was discarded prior to any analyses.

Block Randomization

We used block randomization so that we could both avoid end game effects where participants played differently in the final periods knowing that the game would end, but also have player data from 10 periods within each treatment. To explain this procedure, here is the description provided to participants via the instructional video:

> Interactions with your group will be divided into a series of rounds. Each round has differences from other rounds. The instructions for each round will be provided by video before the round begins. Each round is made up of several periods. The probability of a round ending is 10% for each period. What this means is that on average you will play 10 periods, but you may play many more or many fewer. If a round ends before 10 periods, you will still play through to the 10^{th} period, but only the periods before the ending period will be counted for your payment. You will be informed which period was the last period at the end of the round.

The full script and screenshots from the video are shown in the final section (Experimental Protocol). Sample screenshots for each treatment are shown in the next section (Screenshots from Experiment).

Screenshots from Experiment

Below are screenshots for each version of the game illustrating the Player view, Leader view when playing, and Leader view when making decisions regarding other players. All instructions were provided through pre-recorded videos. The script for these videos along with all screenshots can be found in the Experimental Protocol section.

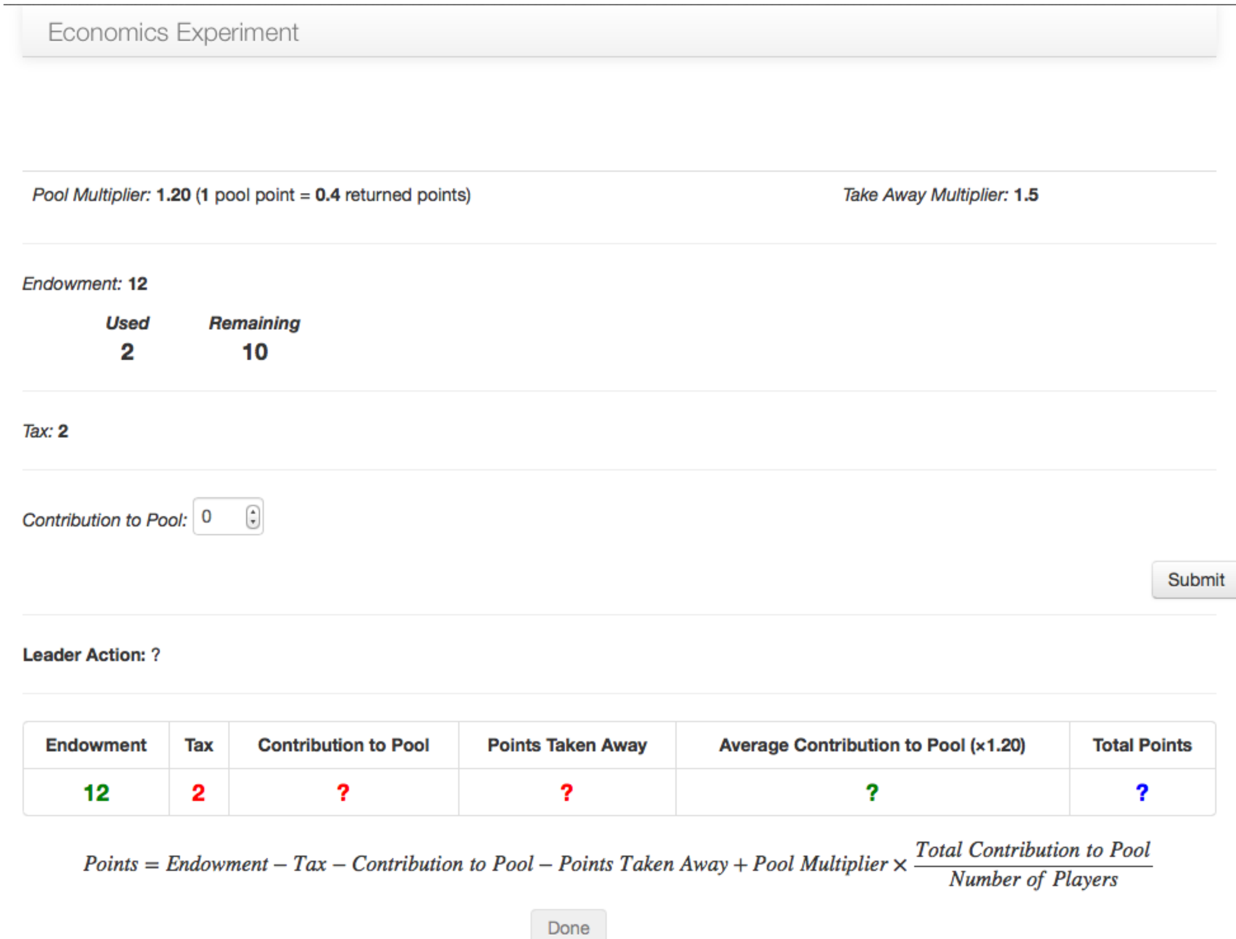


Figure S5. Player screen. The leader's decision (Do Nothing or Take Away Points) is displayed after "Leader Action". It is displayed after the leader has made their choice.

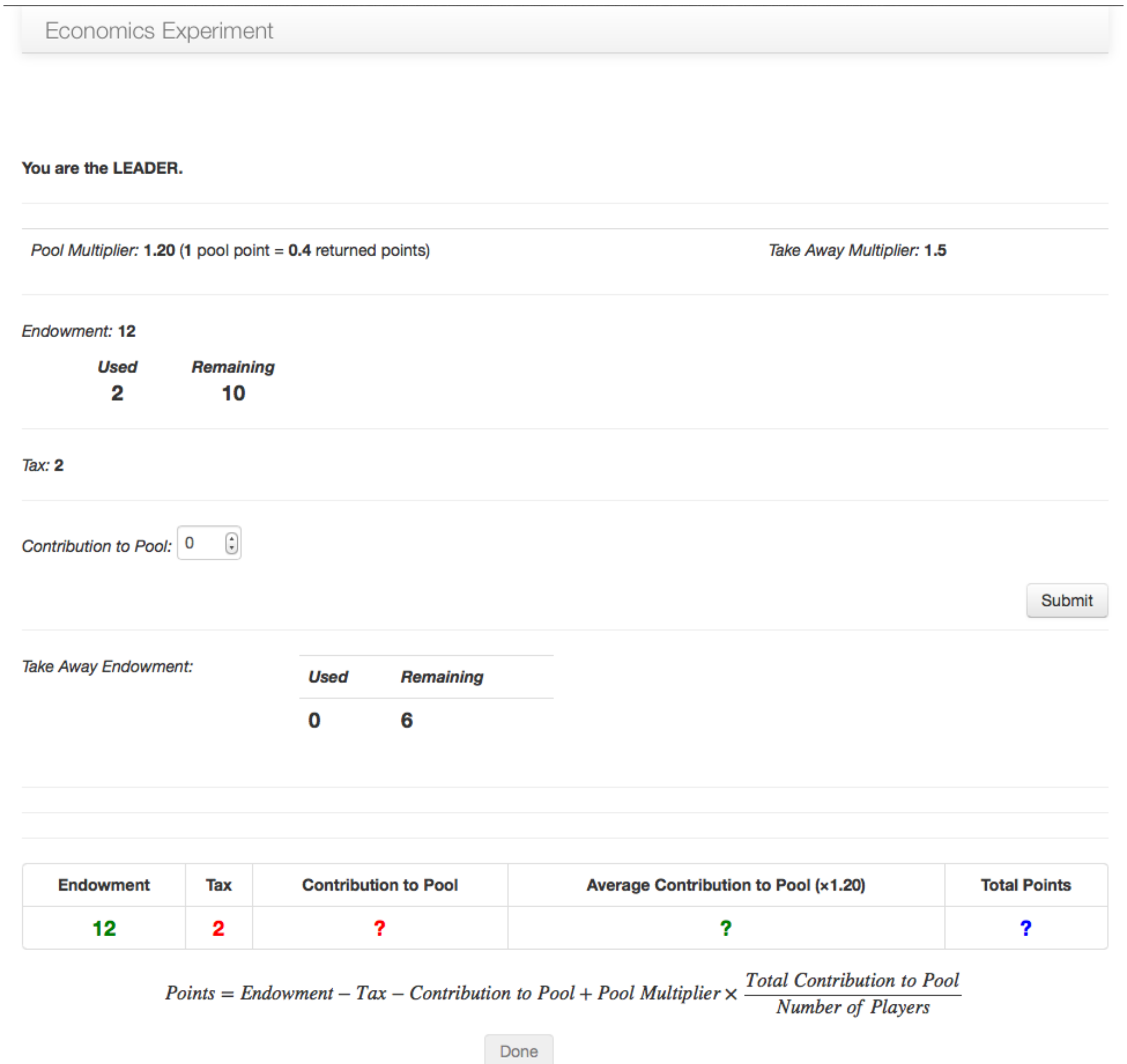


Figure S6. Leader screen for play. After all players have made their decision, leaders can choose how to react to player choices.

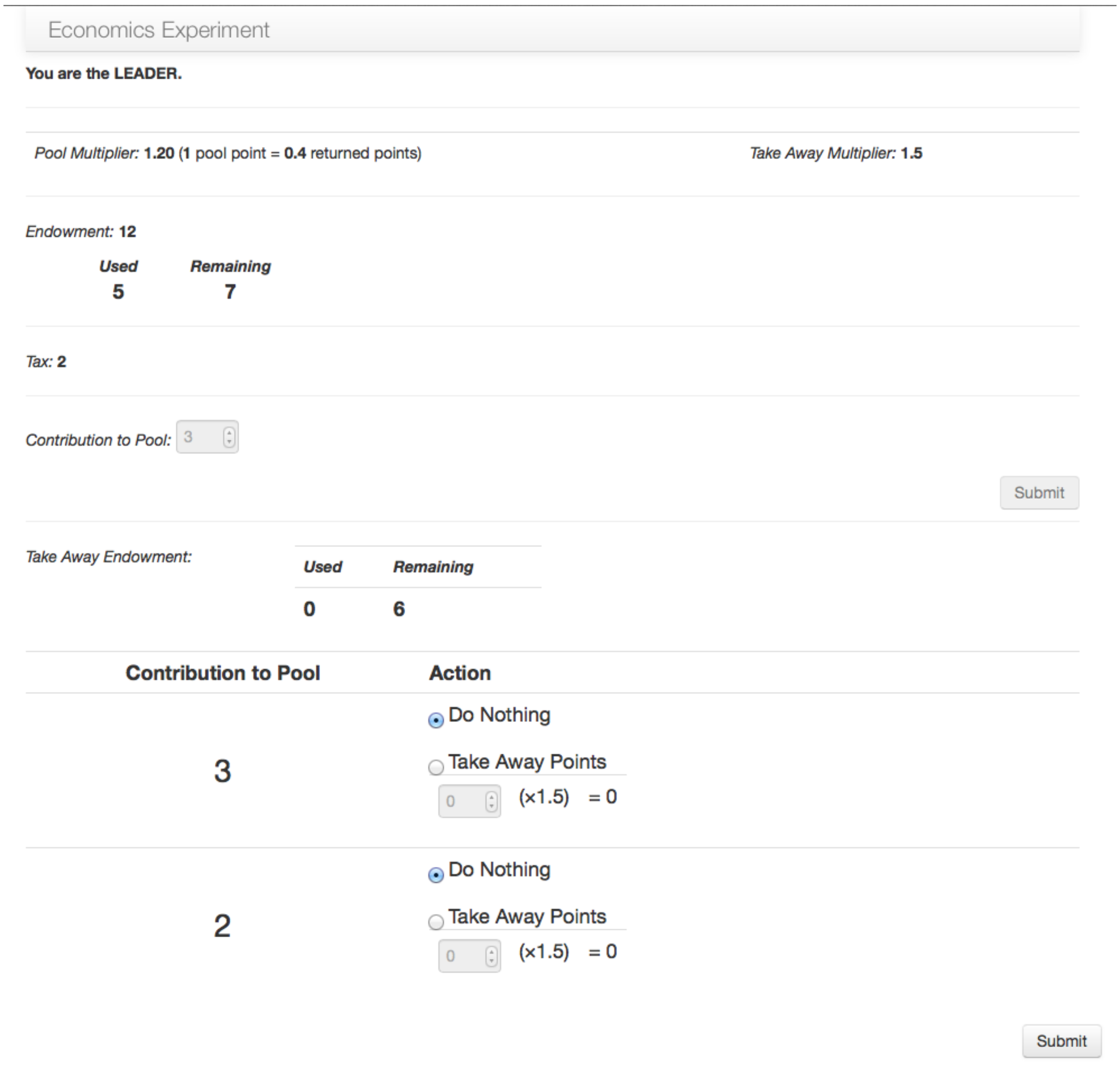


Figure S7. Leader screen for decision regarding players. Leaders are shown anonymized player choices and can choose to Take Points Away or Do Nothing.

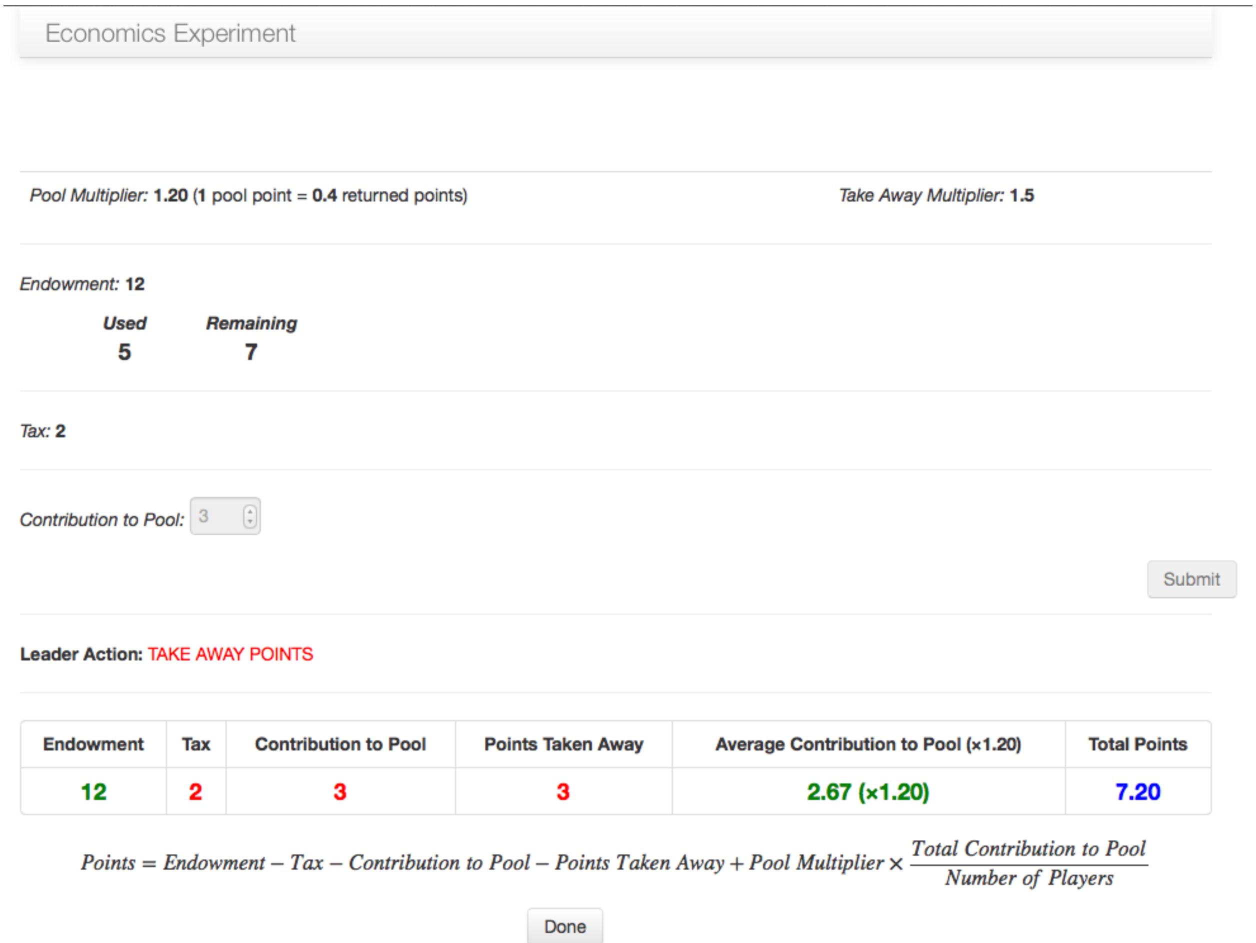


Figure S8. Example of player screen after leader decision. Here the leader has chosen to Take Away Points.

Bribery Game

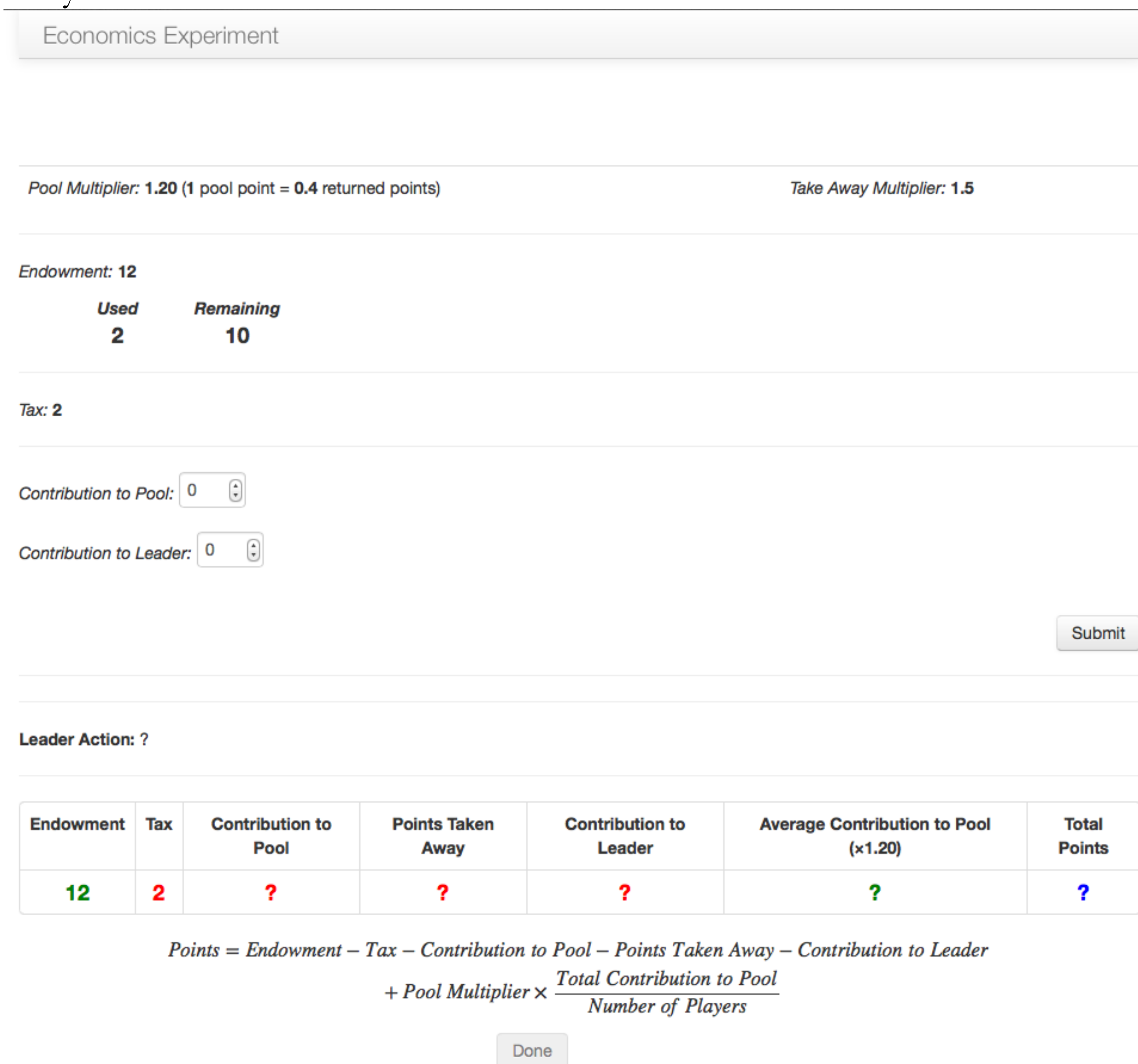


Figure S9. Player screen. Leader action (Do Nothing or Take Away Points) is displayed after the leader has made their decision. The key difference in the Bribery Game is the additional player choice to Contribute to Leader.

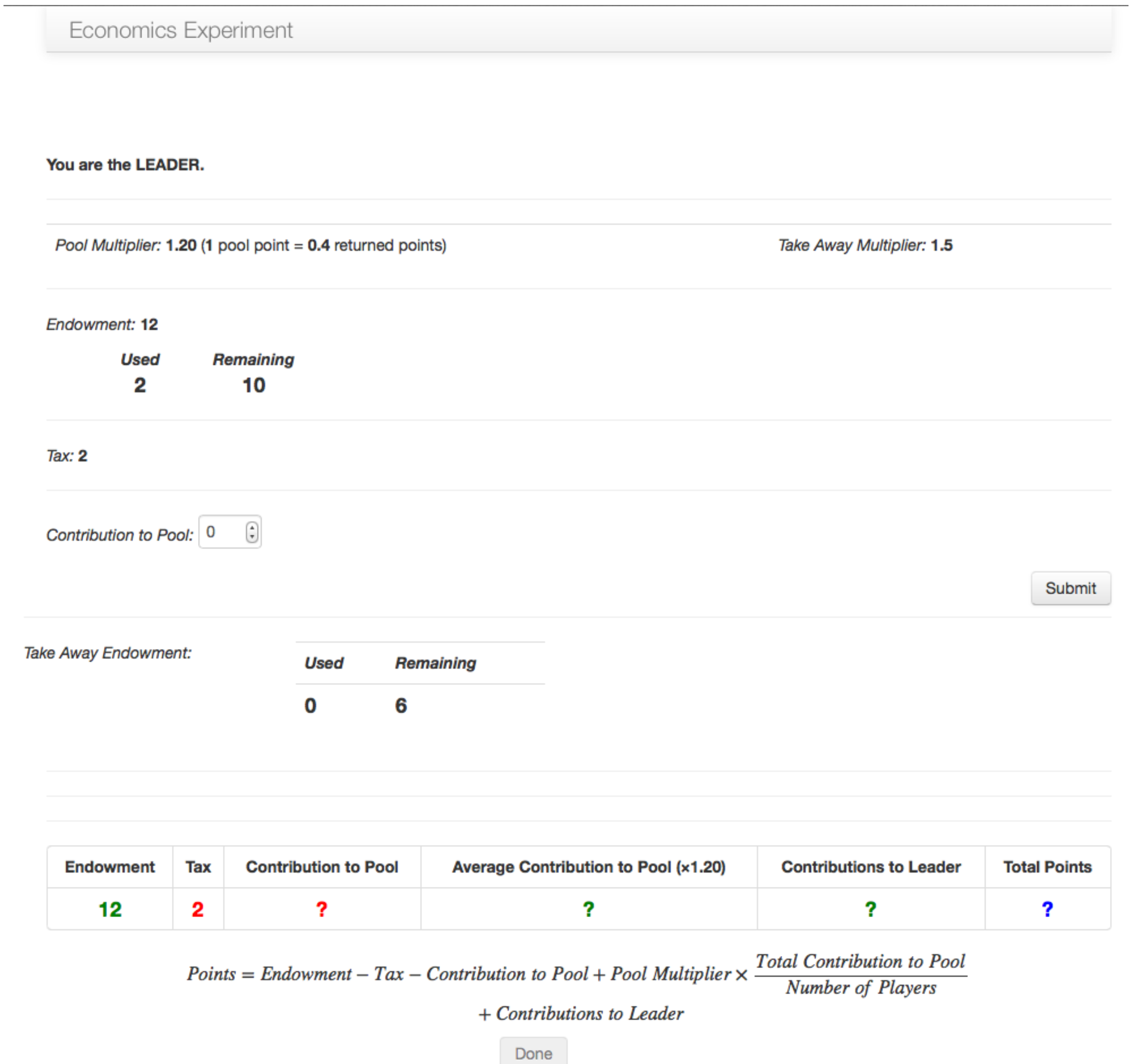


Figure S10. Leader screen for play. Note that leaders cannot contribute to themselves. After all players have made their decision, leaders can choose how to react to player choices.

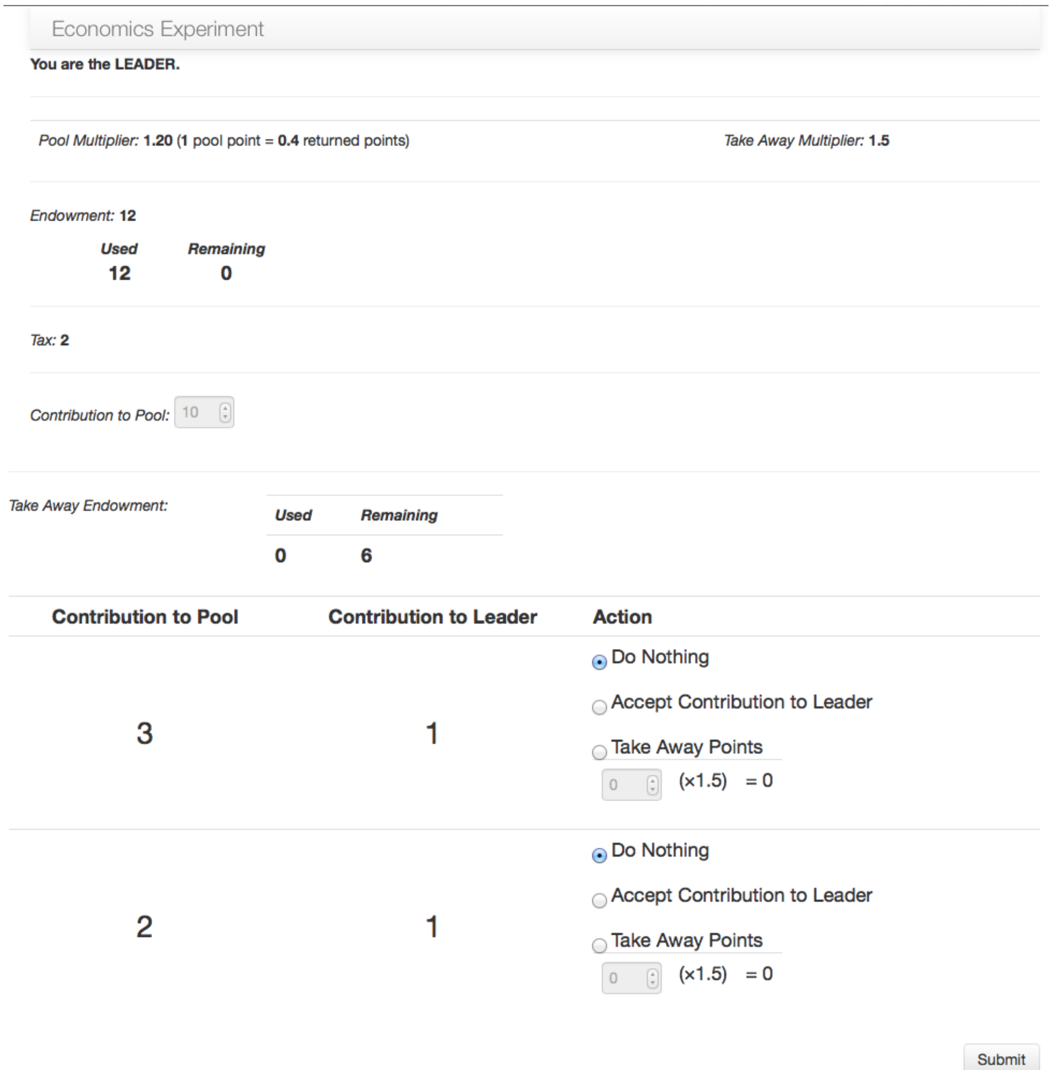


Figure S11. Leader screen for decision regarding players. Leaders are shown anonymized player choices and can choose to Take Points Away, Accept Contribution to Leader, or Do Nothing.

Bribery Game with Partial Transparency

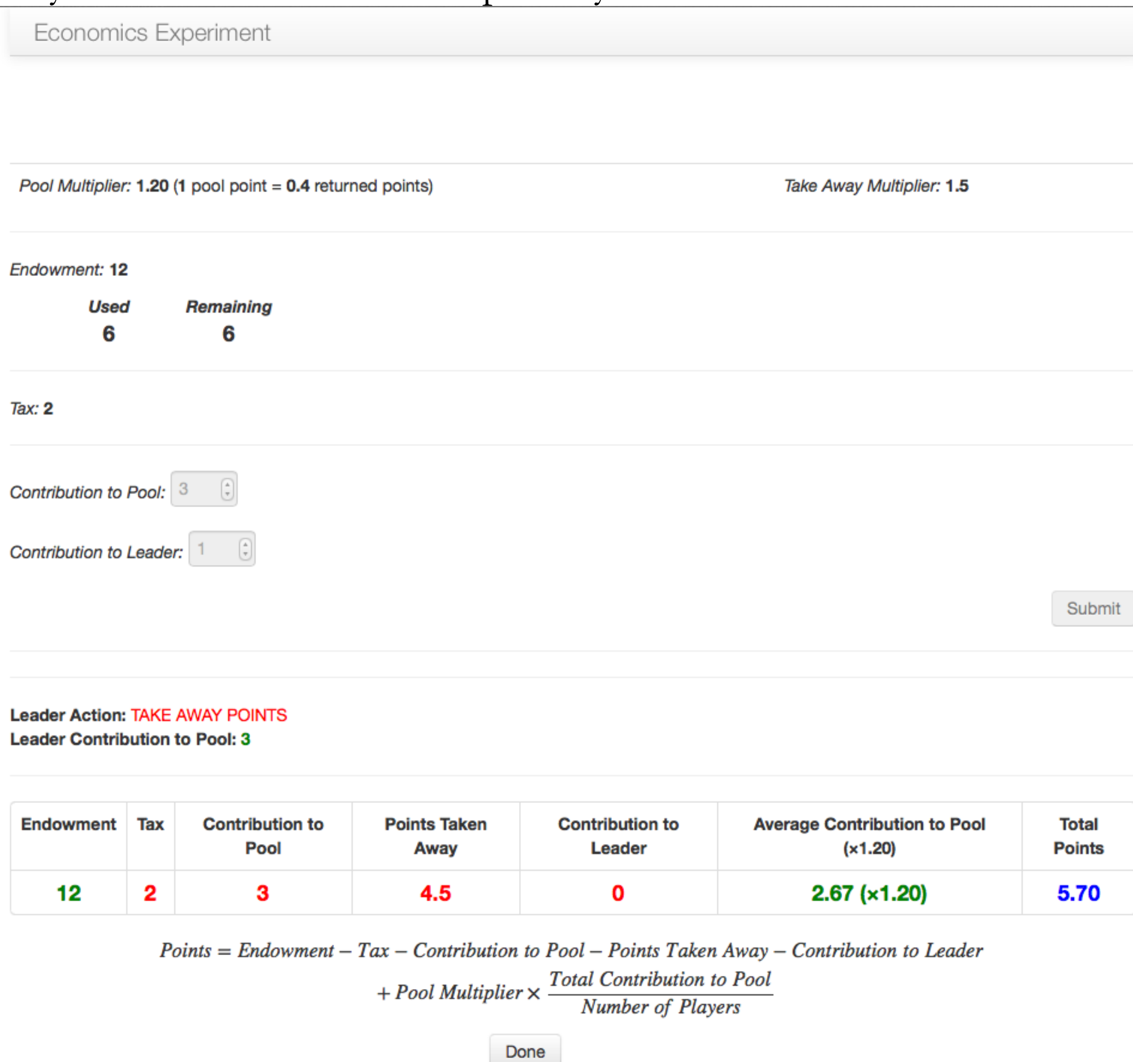
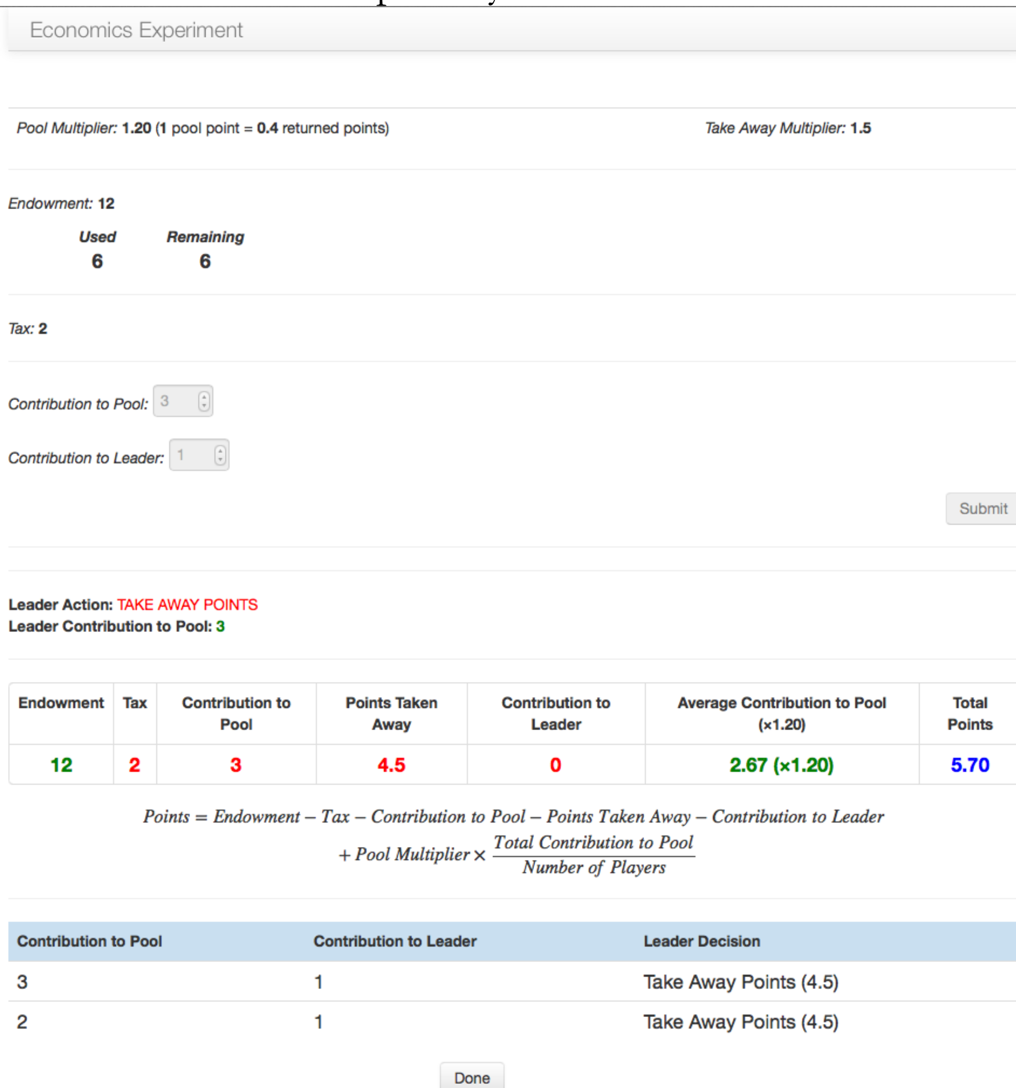


Figure S12. Player screen after Leader decision. Note that below Leader Action (Do Nothing, Take Away Points, or Accept Contribution to Leader) is the Leader's Contribution to the Pool. All other screens are identical to Bribery Game. Here the Leader has contributed 3 points to the public pool and has chosen to take away points from this player.

Bribery Game with Full Transparency



Economics Experiment

Pool Multiplier: **1.20** (**1** pool point = **0.4** returned points) *Take Away Multiplier:* **1.5**

Endowment: **12**

Used	***Remaining***
6	**6**

Tax: **2**

Contribution to Pool: 3

Contribution to Leader: 1

Submit

Leader Action: TAKE AWAY POINTS
Leader Contribution to Pool: 3

Endowment	Tax	Contribution to Pool	Points Taken Away	Contribution to Leader	Average Contribution to Pool (×1.20)	Total Points
12	2	3	4.5	0	2.67 (×1.20)	5.70

$$Points = Endowment - Tax - Contribution\ to\ Pool - Points\ Taken\ Away - Contribution\ to\ Leader + Pool\ Multiplier \times \frac{Total\ Contribution\ to\ Pool}{Number\ of\ Players}$$

Contribution to Pool	Contribution to Leader	Leader Decision
3	1	Take Away Points (4.5)
2	1	Take Away Points (4.5)

Done

Figure S13. Player screen after Leader decision. Note that below Leader Action (Do Nothing, Take Away Points, or Accept Contribution to Leader) is the Leader's Contribution to the Pool, as in Bribery Game with Partial Transparency. However, now all Leader and Player Actions are displayed in a table. All other screens are identical to Bribery Game.

Bribery Game with Forced Leader Contribution

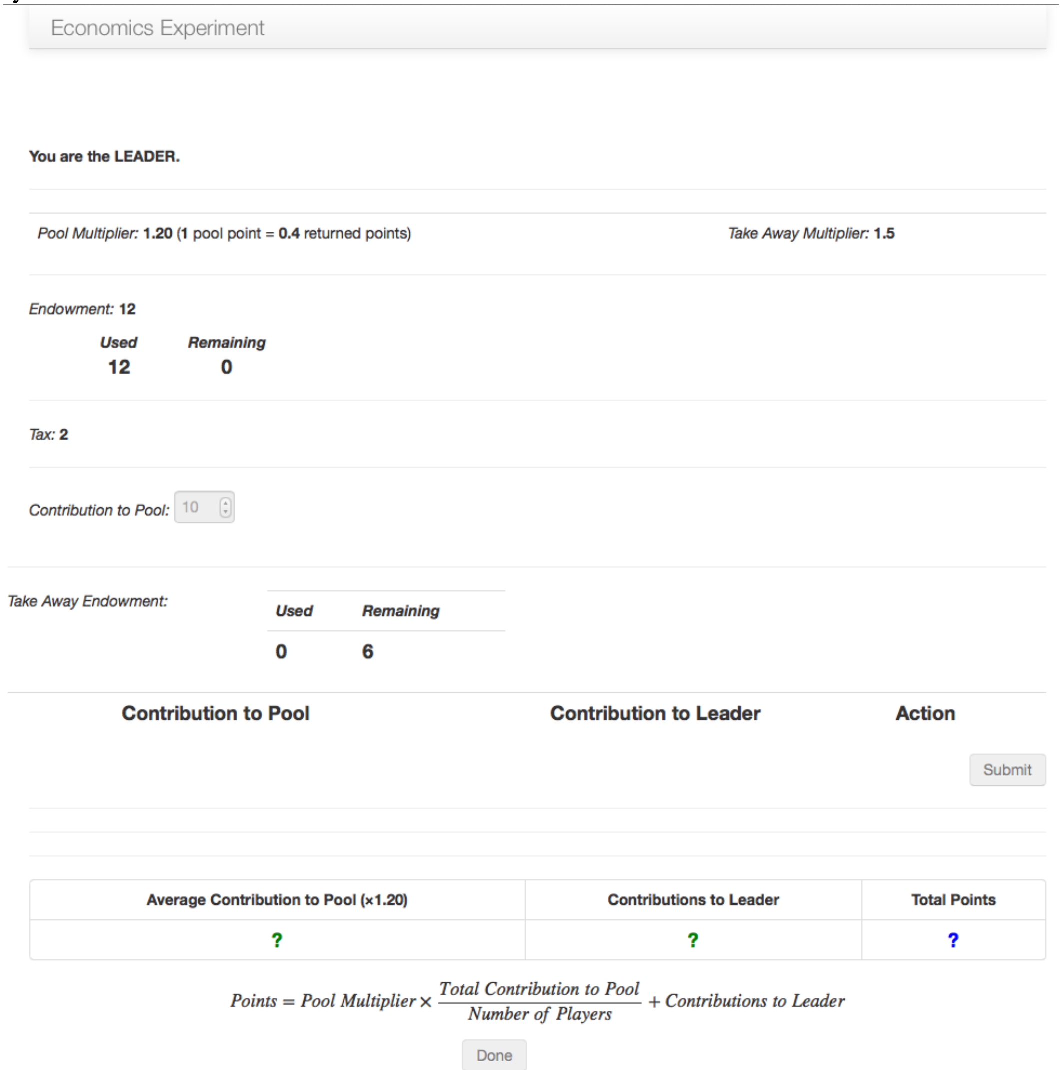


Figure S14. Leader screen for play. Note that all leader points are automatically contributed to the pool. All other screens are identical to Bribery Game.

Experimental Protocol

The videos were created using Microsoft Powerpoint and exported as a video. For sections of the instructions that were shared between treatments, the same slide and recording were used to ensure there were not even minor differences in aspects as such intonation. The following is a transcript of the experimental video with screenshots. All participants first heard the General Description. They then heard the First Round version of the treatment they received first. All subsequent treatments had the Subquent Round version of the treatment instructions, which highlighted the difference with other treatments. Also included are the quiz questions administered to all participants. These quizzes were used to ensure participants understood the instructions and the differences between treatments. Participants had to answer all questions correctly before they could proceed.

The instructional videos infomed participants that they would receive 10c per point and be paid for 8 random rounds. However, early pilots testing the software suggested that this would not result in adequate conpensation for participants as per <Institution Subject Pools> policies. Rather than record all instructional videos, we instead informed participants that they would actually be paid 15c per point for 10 random rounds rather than 10c per point for 8 random rounds, but that all other aspects of the instructions were correct and unchanged. No participants had questions or complained about this change.

General Description

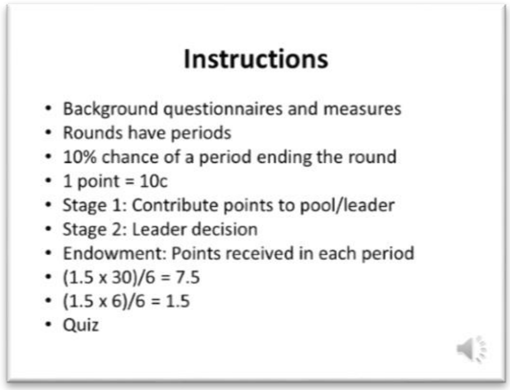
Welcome to the experiment. All instructions will be provided through videos. It is prohibited to communicate with other participants during this experiment, however, please feel free to ask the experimenter any questions.

You are invited to participate in this experiment. When you are ready, please pause the video to read and sign the consent form.

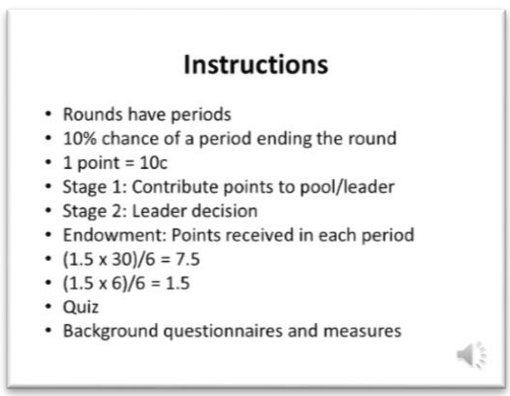
<PAUSE>

If you are happy to continue, we may now begin the study. First, please turn off your mobile phone or other electronic devices.

There is no deception in this experiment: all other group members are real people and all monetary amounts are paid precisely as described. If you need further clarification about any aspect of this experiment, please let the experimenter know before we begin.



[VERSION A: You will begin by completing some background questionnaires and measures. You will then engage in a series of interactions with a group of other people via the computer and will be asked to make some decisions.]



[VERSION B: You will engage in a series of interactions with a group of other people via the computer and will be asked to make some decisions.]

Interactions with your group will be divided into a series of rounds. Each round has differences from other rounds. The instructions for each round will be provided by video before the round begins. Each round is made up of several periods. The probability of a round ending is 10% for each period. What this means is that on average you will play 10 periods, but you may play many more or many fewer. If a round ends before 10 periods, you will still play through to the 10th period, but only the periods before the ending period will be counted for your payment. You will be informed which period was the last period at the end of the round.

During this experiment, we will refer to points. Every point is worth 10 cents in today's experiment. You will be paid for 8 periods randomly selected from across all rounds. Together with your $10 show up fee, you could earn over $35 today, but the precise amount depends on your actions and the actions of other group members.

In each period the experiment consists of two stages. At the first stage you have to decide how many points you would like to contribute to a project, refered to as a pool, and to a group member designated as a leader. In the second stage, the group member designated as a leader will make a decision whether to accept points from group members or whether or how much to reduce the earnings of group members from the first stage. The leader is also a group member in the first stage, but does not make a decision about themselves.

At the beginning of each period each participant receives an endowment of points. We call this his or her endowment. Your task is to decide how to use your endowment. You have to decide how many of the points you want to contribute to the pool, how many you want to contribute to the leader, and how many of them to keep for yourself. All your decisions are anonymous.

The income of each group member from the pool is calculated in the same way, this means that each group member receives the same income from the pool, regardless of how much they contributed. Suppose the pool multiplier is 1.5 and the sum of the contributions of all group members is 30 points. In this case each member of the group receives an income from the project of: 1.5 times 30 = 45 divided by the number of group members. So if there are 6 group members, each group member would receive 45 divided by 6, or 7.5 points. If the total contribution to the project is 6 points, then each member of the group receives an income of 1.5 times 6 = 9 divided by the number of group members. So for 6 group members, each group member would receive 9 divided by 6, or 1.5 points.

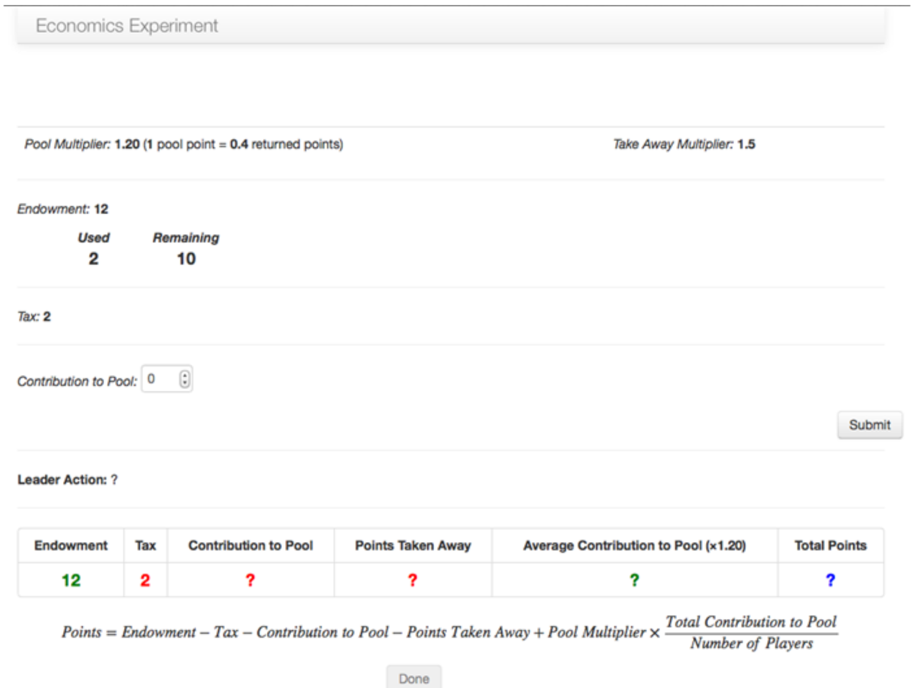
Each point you keep for yourself remains as 1 point. Any points given to the leader, if accepted are kept by the leader. If not accepted by the leader, these points are returned to you. You will be informed of the number of group members after these instructions.

The instructions for each round will be provided before the round begins. You will be given a quiz after the video, which you must answer correctly to continue, so please pay attention. When you are ready, you may watch the first video. When everyone has watched the video and passed the quiz and there are no further questions, the experiment will begin.

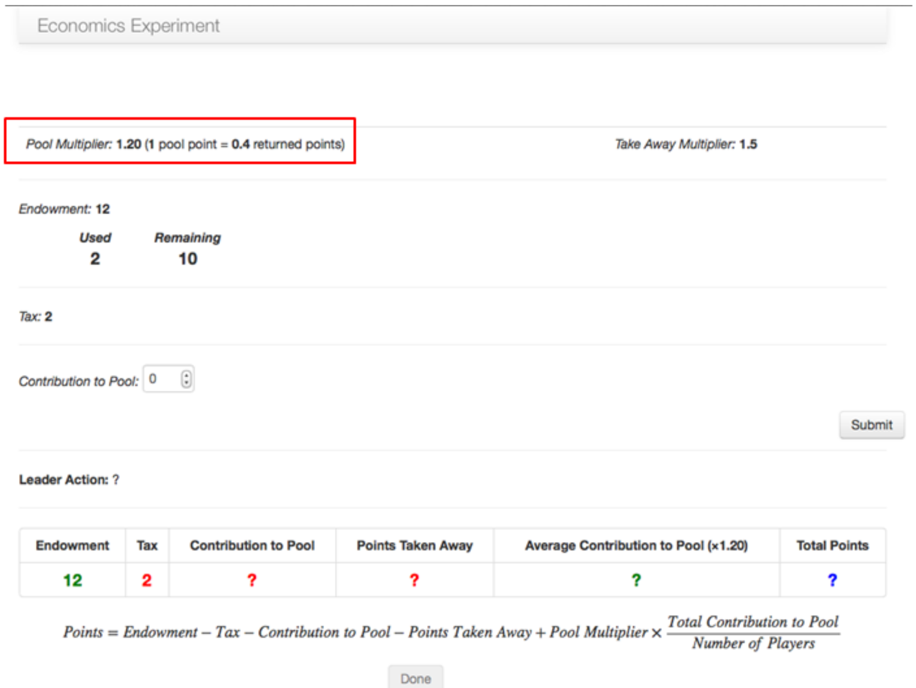
[VERSION B: After all games are completed, you will complete some background questionnaires and measures.]

Control Treatment

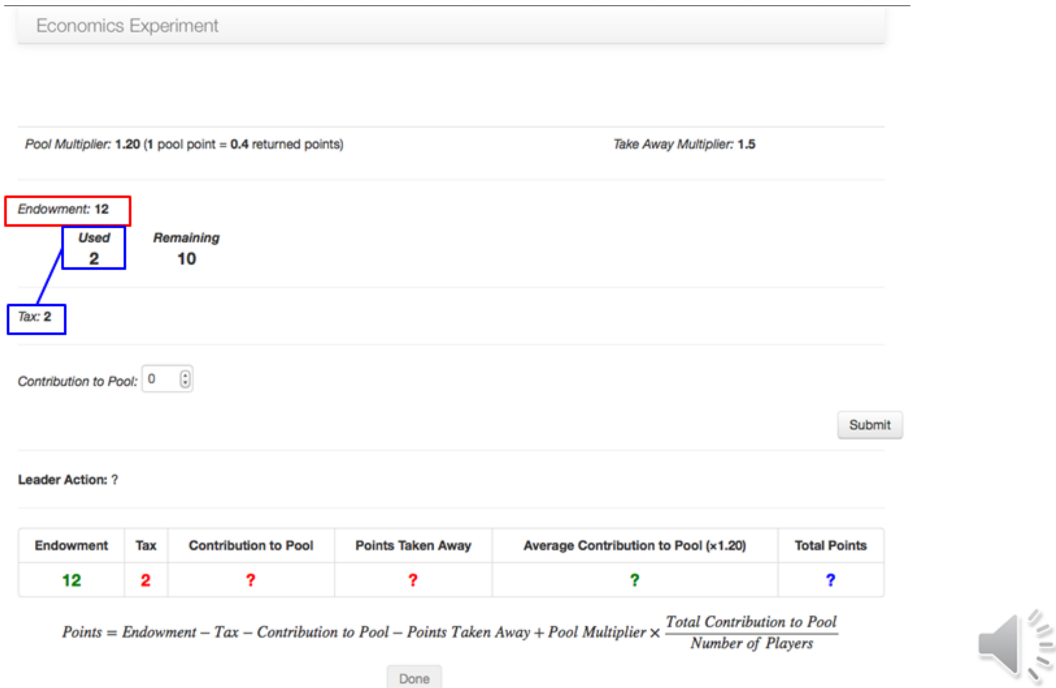
First Round



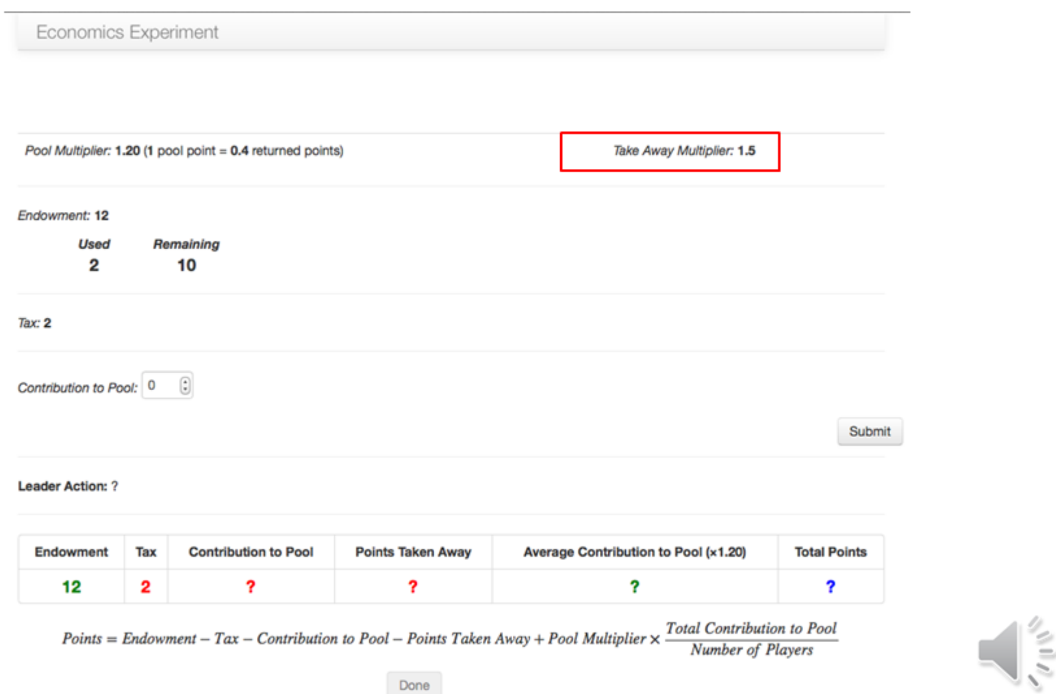
Welcome to the first round. The screen you see in front of you is the normal screen for play. In each period, one person from the group will be selected at random to play the role of a leader. The leader will have a slightly different screen. Please pay attention to all screens presented in these instructions as the instructions for other other rounds will only highlight the differences between that round and this one.



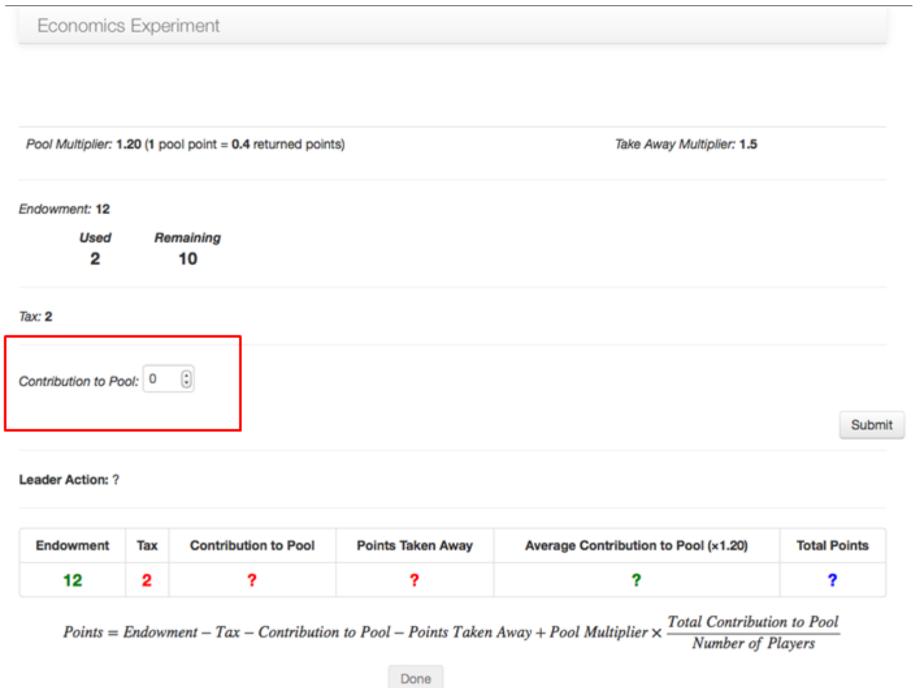
The first thing we'd like to draw your attention to is the pool multiplier. When group members allocate points to the group pool, the total allocated by the group is multipled by this number, the pool multiplier. In this case, it's 1.2. The new total is then distributed equally among everyone in the group, including the leader. This means that if every group member allocated all their points, all group members would increase their earnings by 20%. If no one allocates to the group pool, there's nothing to multiply by 1.2, and nothing to divide up. Another way to think about this is to consider how much each point in the pool is worth to each player. This is shown within parantheses to the right of the pool multiplier. In this case, with 3 players, each player gets 0.4 points for every point in the pool, no matter who contributes it. For example, if each player contributed 1 point pool, there would be 3 points in the pool and each player would get back 3 times 0.4 or 1.2 points.



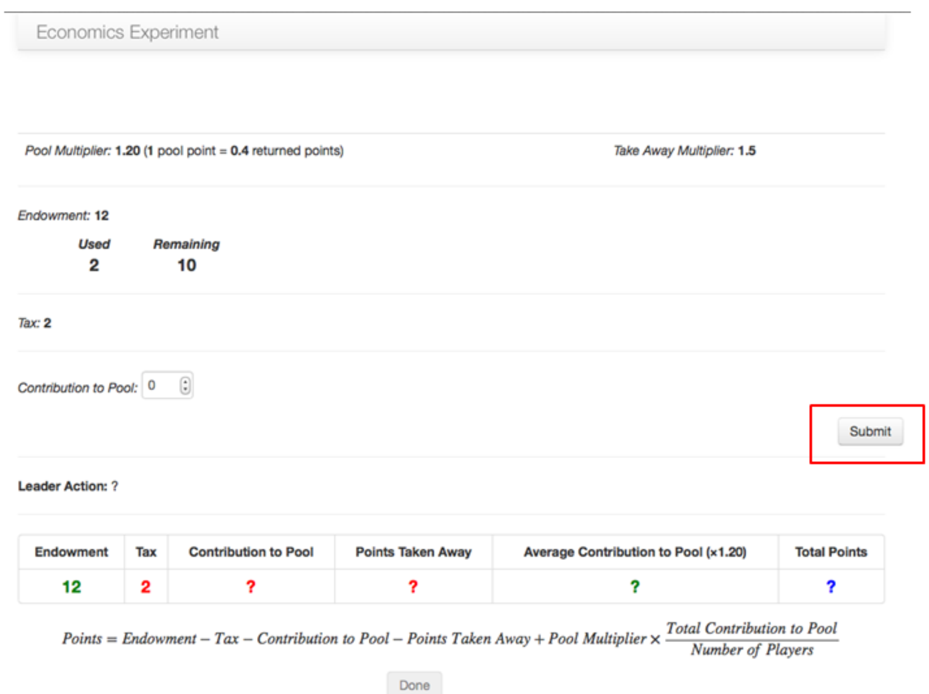
Below this, we see your initial endowment of 12 points. 2 points are taxed each period, which can be used by the leader to take points away from group members. This leaves you 10 points to allocate.



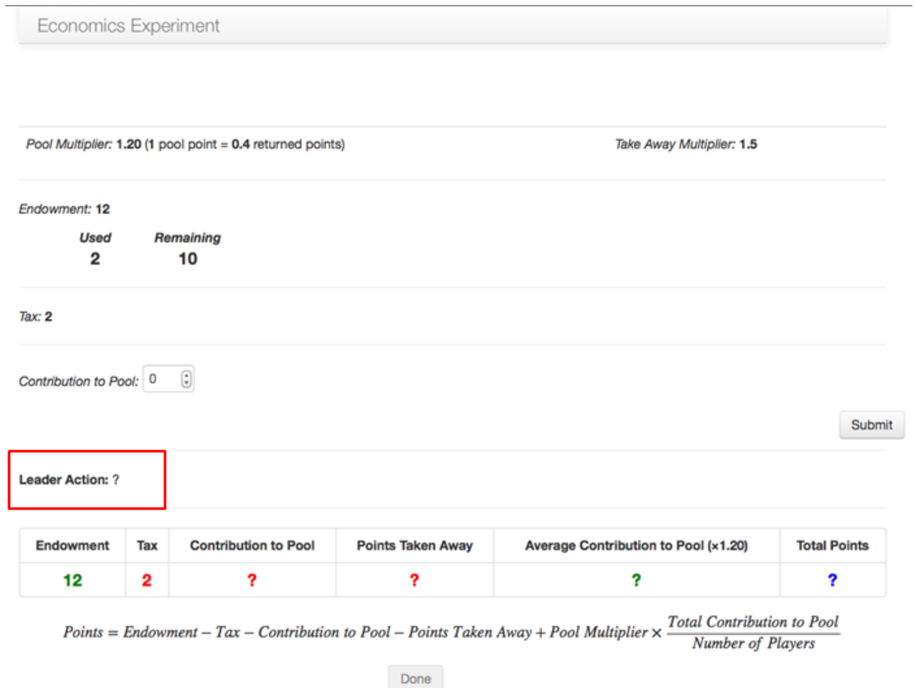
Here you see the Take Away multiplier. This is how much each tax point is multiplied by when the leader takes points away from a group member. For example, here, if the leader spent 1 tax point to take points away, 1.5 points would be taken from the targeted group member.



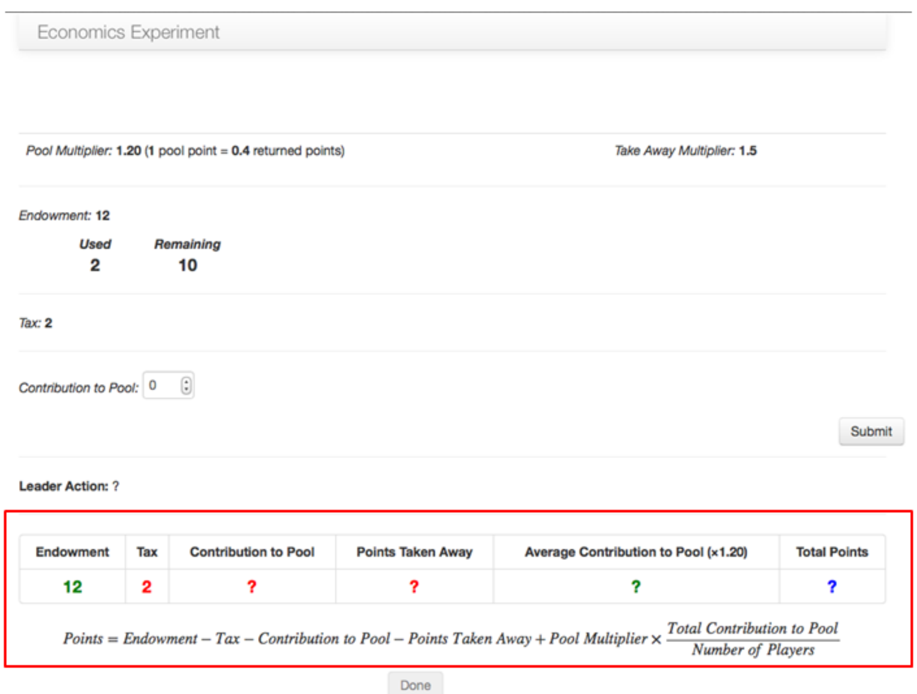
Here we see the boxes that are used to allocate your remaining 10 points to the group pool. Any points you do not allocate will be kept for yourself. When you hover your mouse over these boxes you will see an arrow to increase or decrease the amount in 1 point intervals or you may type in a value.



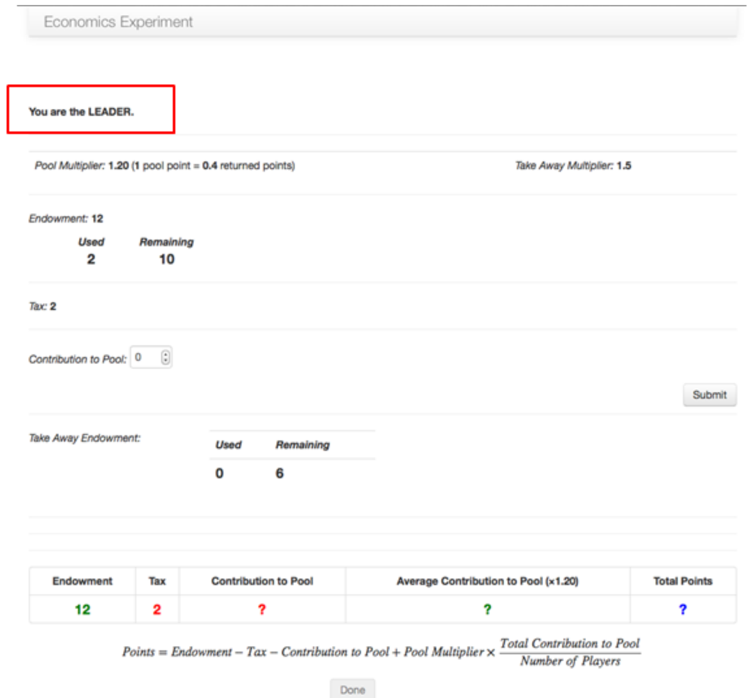
When you have made your decision, click “Submit”. After clicking “Submit”, you will not be able to change your decision.



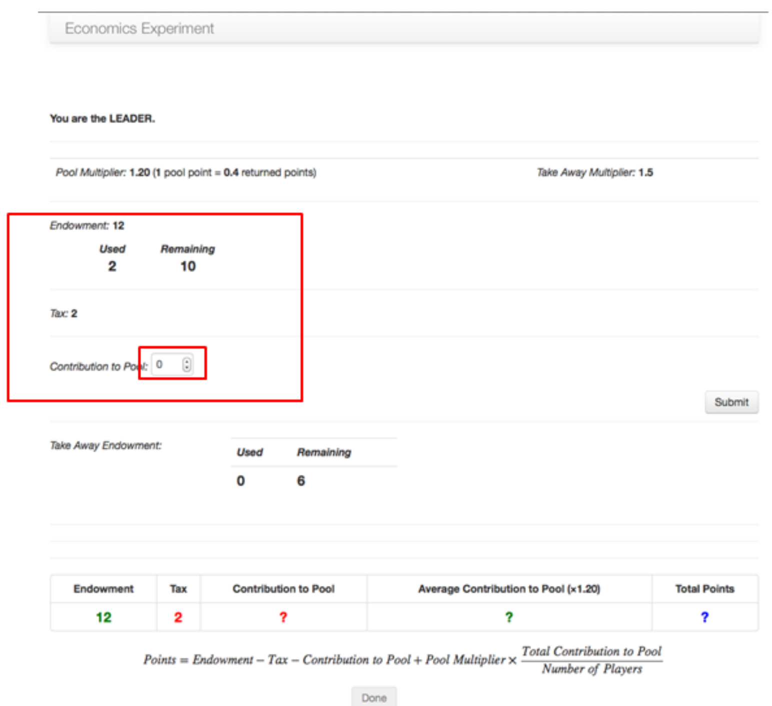
This section will show you the leader's action. The leader has 2 choices – to take points from you or do nothing. They may only choose one of these options.



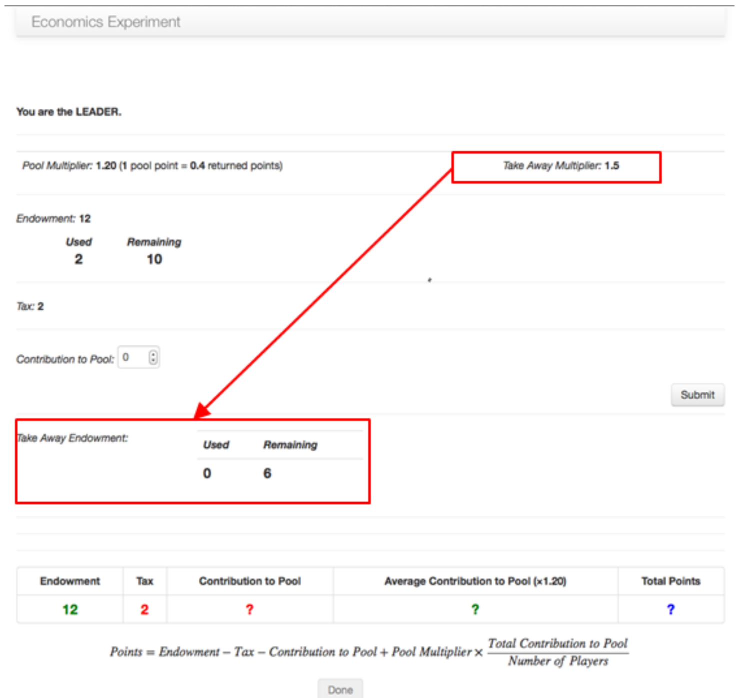
Here you see how your points for this period are calculated. Everything in green is added, everything in red is subtracted, and your total points for the period is shown in blue on the right. We will return to this after explaining the leader's screen.



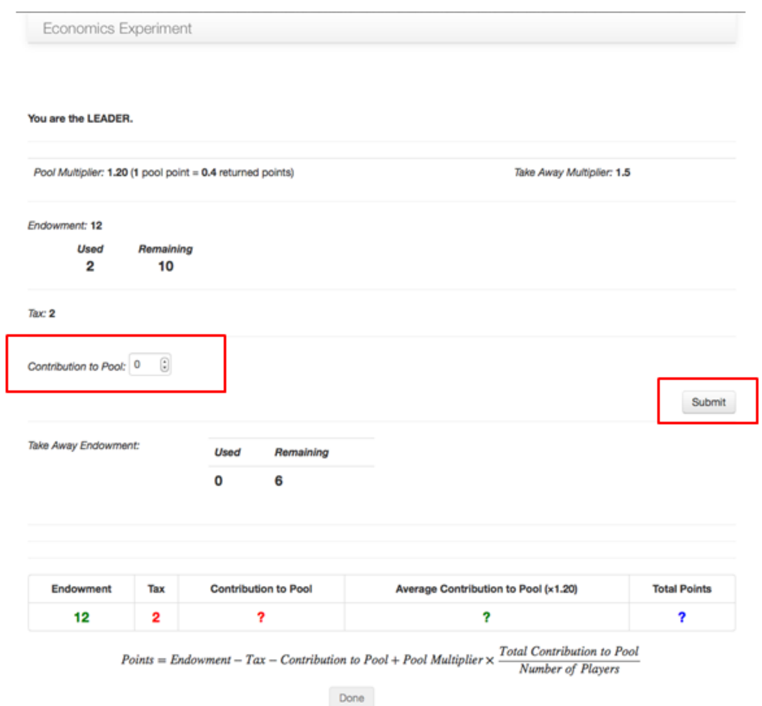
If you are the leader for a period, you will be identified at the top of the screen. Note that all group members, including the leader are anonymous to other group members. Your screen looks very similar to other group members, with a few differences.



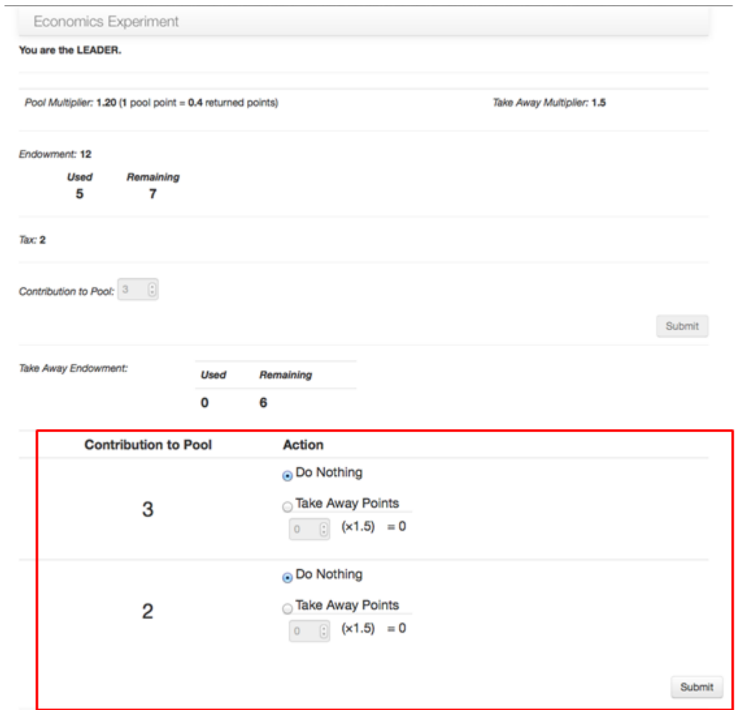
Like other group members you pay a tax and choose how much you wish to allocate to the group pool and how much to keep for yourself



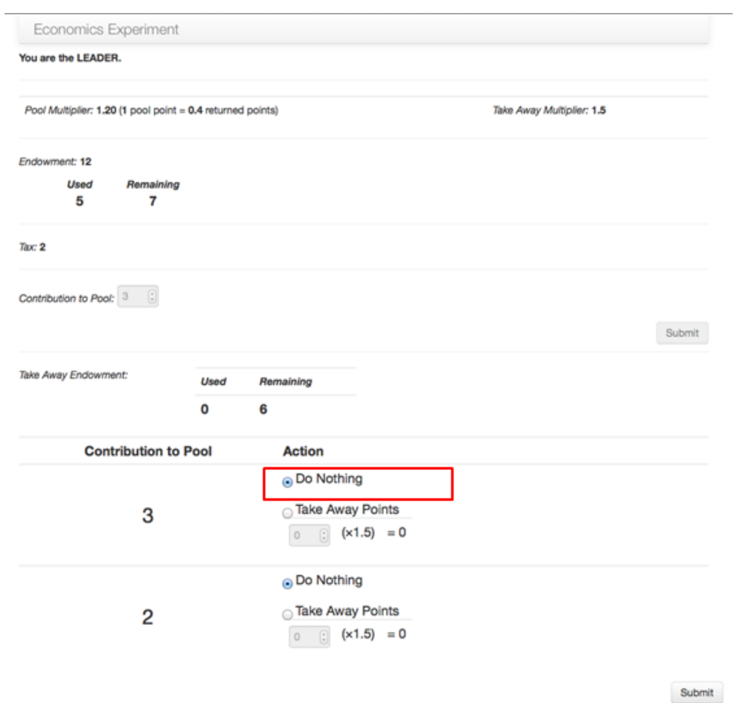
You also have access to tax points that you may use for taking away points from group members based on the Take Away multiplier.



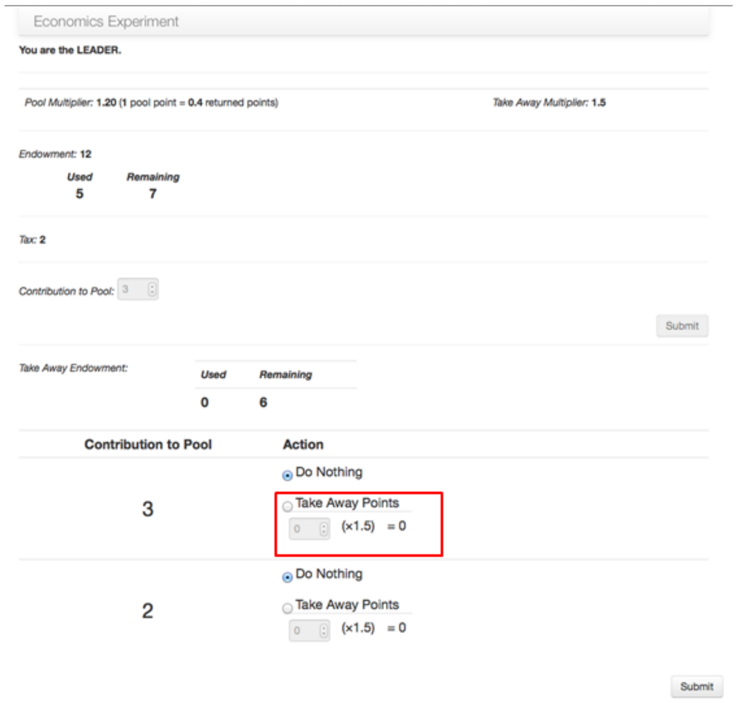
The actions of group members will appear below this after they make their decision. The Contribution column are the contributions to the group pool.



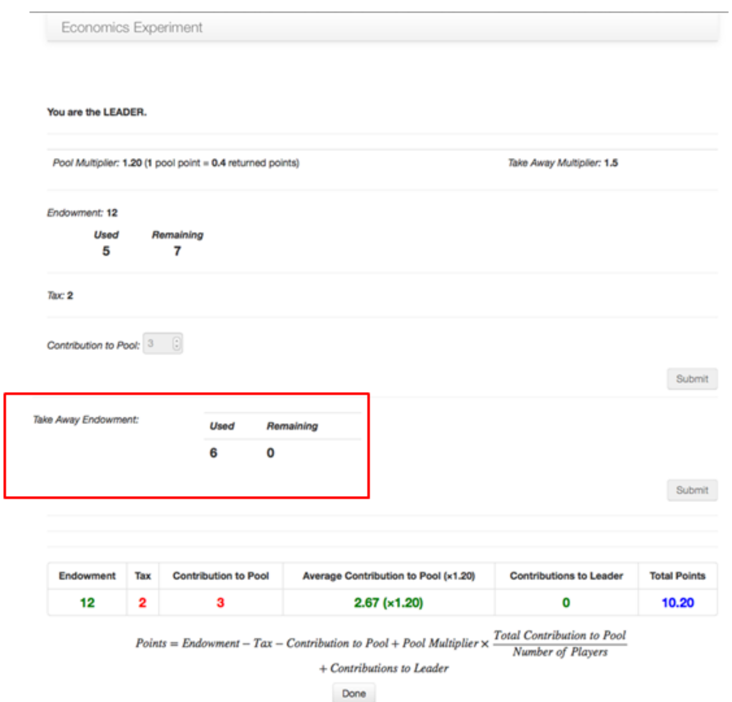
You may…



do nothing…

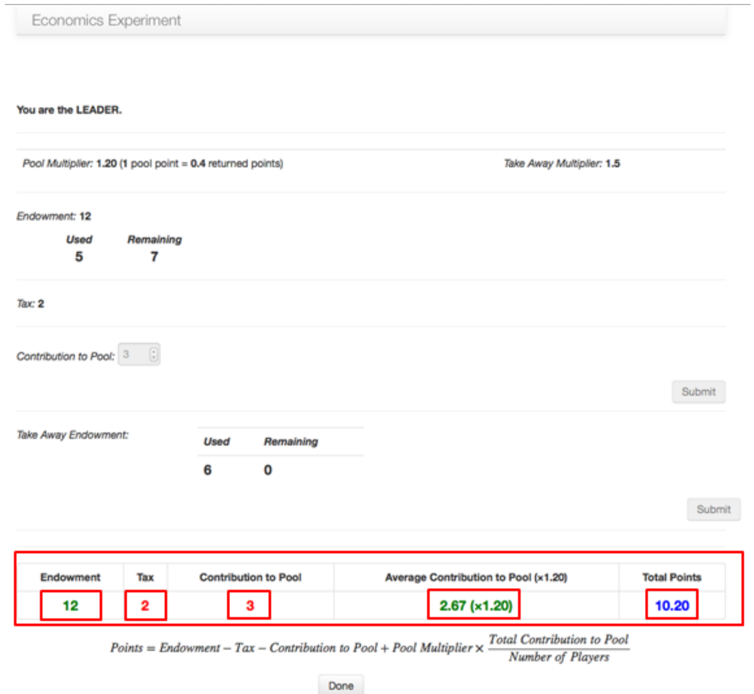


or take points away from any group member. If you take away points, you decide how much to take by allocating tax points to the group member. The total points taken away is calculated for you next to this box.

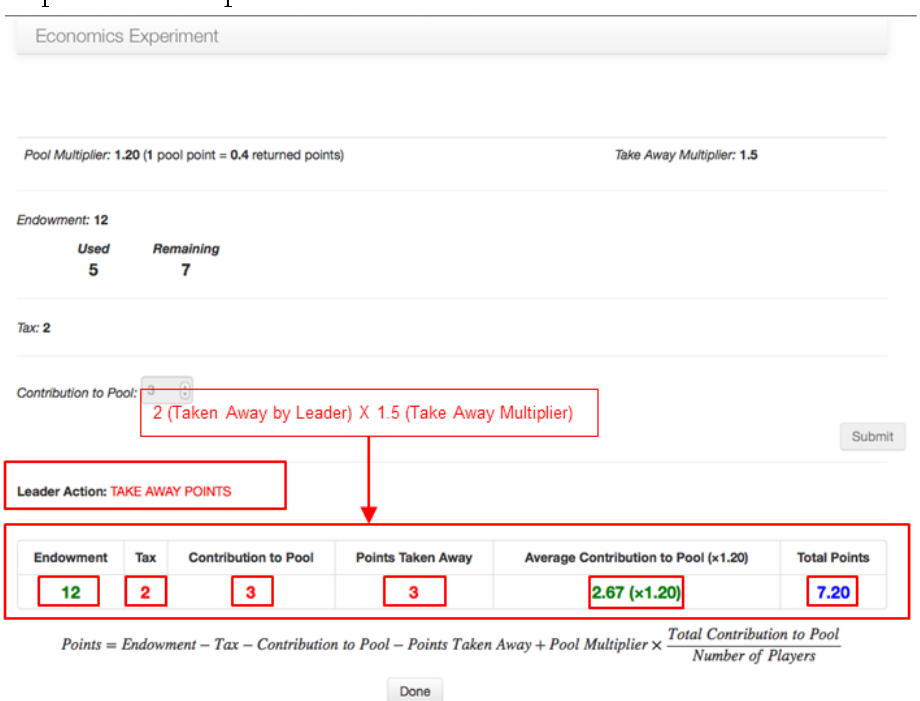


Any tax points not used for taking away points from group members will be returned to the experimenter. Similarly any points taken away from group members will also be returned to the experimenter.

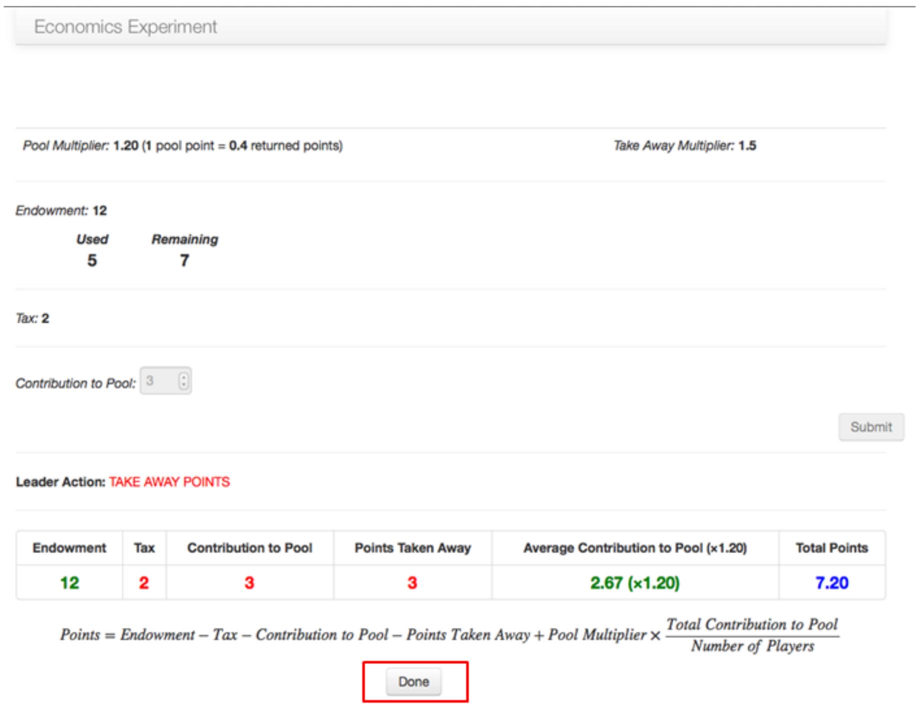
As a leader your earnings for this period are determined by how much you receive from your share of the group pool.



This is your endowment minus your taxes minus your contribution to the pool. The average contribution of group members is shown here – 2.67 points. This is multiplied by the pool multiplier to give the total points for this period – 10.20.



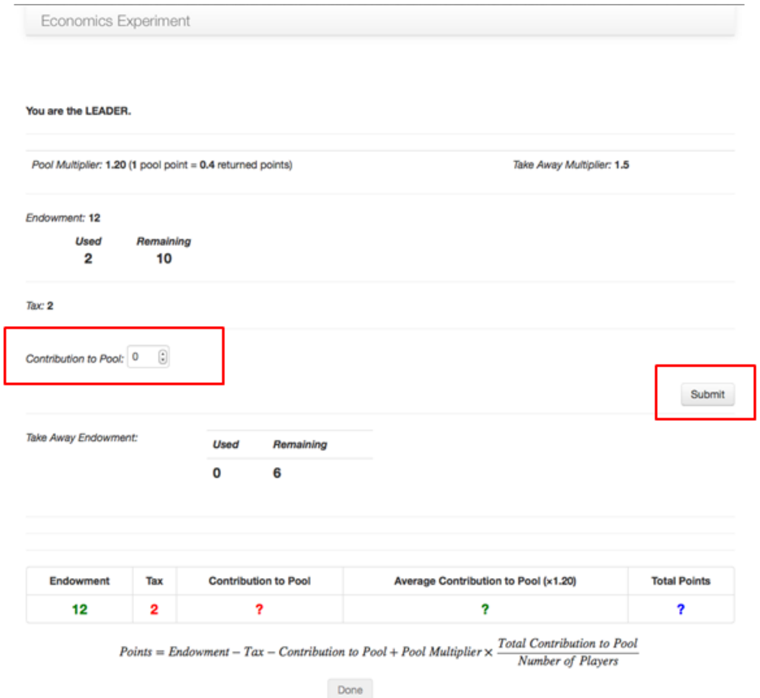
After the leader clicks submit, group members will be informed of the leader's action. The earnings of group members for this period will be calculated from their initial endowment minus their taxes, minus their contribution. Then minus the points taken away by the leader, if any. In the example here, the leader chose to take points away from the group member. Finally, the group members share of the group pool will be added to this amount to calculate the total points for this period. Here the average contribution is shown – 2.67 points. This is multiplied by the pool multiplier and then added to give total points – 7.20 for this group member.



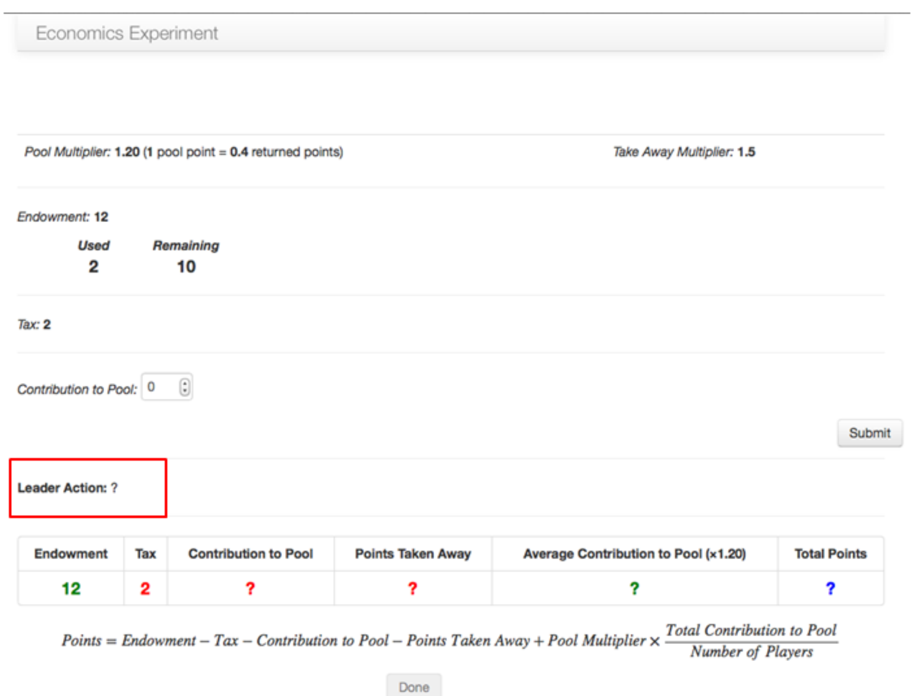
Please click "Done" at the bottom of the screen to move to the next period.

Please click "Continue" when you are ready to begin.

Subsequent Round



The difference between this round and previous rounds is that in this round, like other group members, the leader must choose how much to contribute to the group pool and can not accept any points.



In addition, group members only see the leader's action towards them. They do not see the actions

of others group members or the leader's actions towards other group members. As in all rounds, all participants are completely anonymous.

Please click "Continue" when you are ready to begin.

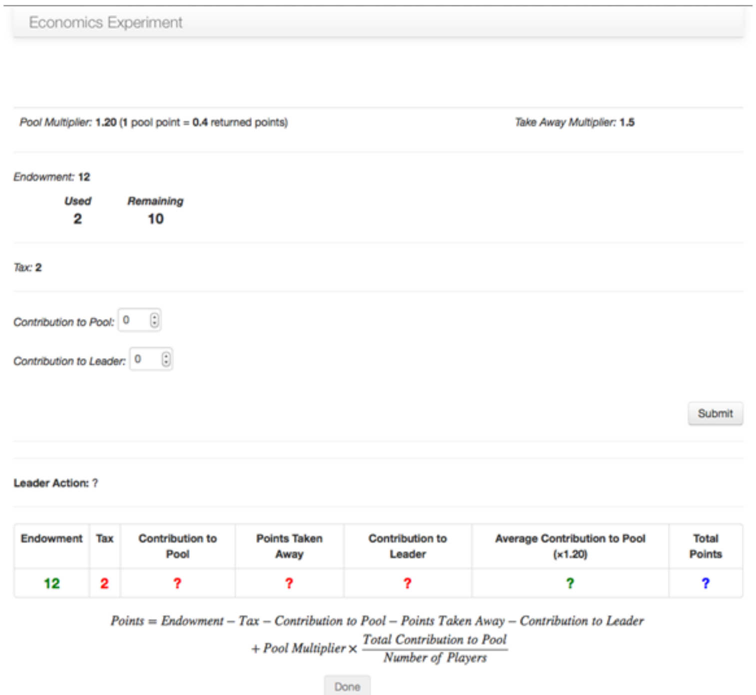
Quiz Questions

(The correct answer is underlined)

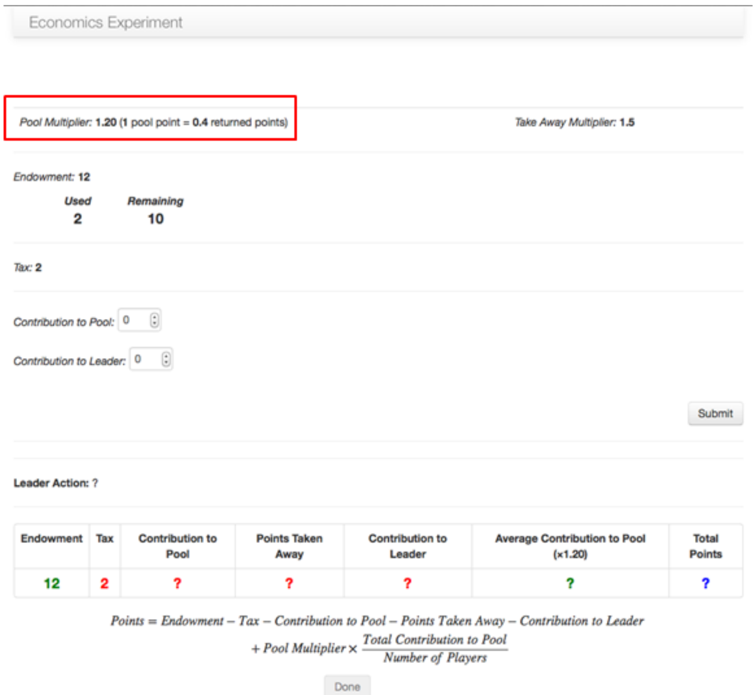
1. How many periods will you play?
    a. 10 periods
    b. More than 10 periods
    c. Less than 10 periods
    d. <u>An average of 10 periods, but possibly more and possibly less</u>
2. If everyone contributes the same amount to the group pool, will you:
    a. <u>Increase your points from your initial endowment</u>
    b. Decrease your points from your initial endowment
    c. Have the same points as your initial endowment
    d. Might increase your points or might decrease your points
3. How much does the leader contribute to the group pool?
    a. <u>However much they choose to allocate to the group pool</u>
    b. 10 points
    c. Depends on the points given to the leader
    d. Depends on the group member's contributions
4. Which of the following can the leader NOT do?
    a. Nothing
    b. <u>Accept points from a group member</u>
    c. Take points away from a group member
    d. Accept points and take points away from the same group member
5. What actions of other group members can group members (not leader) see?
    a. <u>Leader's action towards them</u>
    b. Leader's contribution
    c. Leader's contribution and all actions
    d. All group member and leader actions

Bribery Game

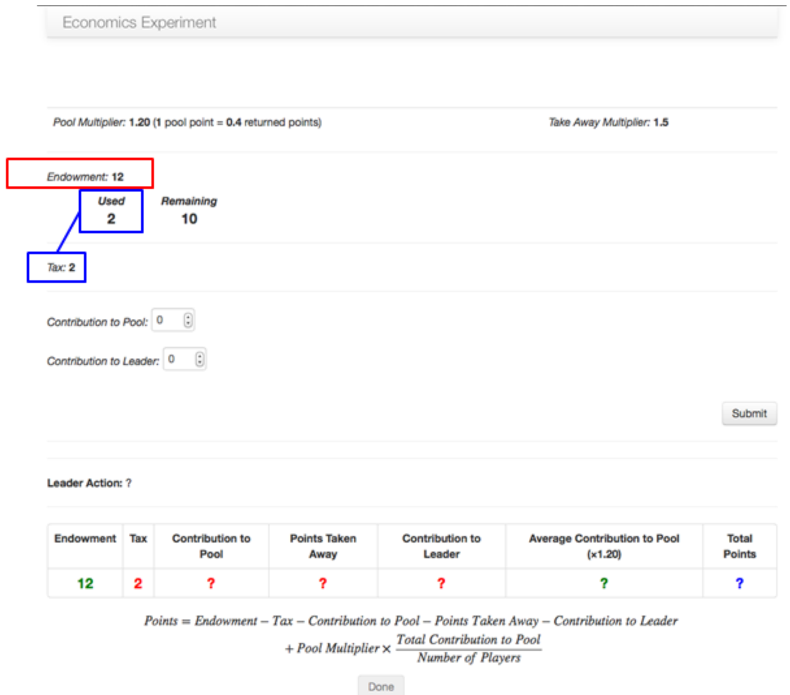
First Round



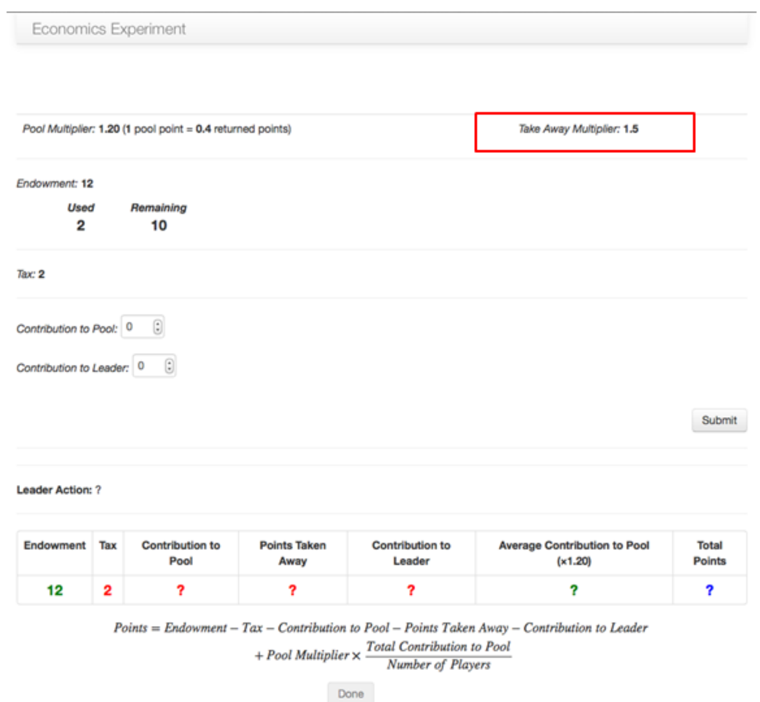
Welcome to the first round. The screen you see in front of you is the normal screen for play. In each period, one person from the group will be selected at random to play the role of a leader. The leader will have a slightly different screen. Please pay attention to all screens presented in these instructions as the instructions for other other rounds will only highlight the differences between that round and this one.



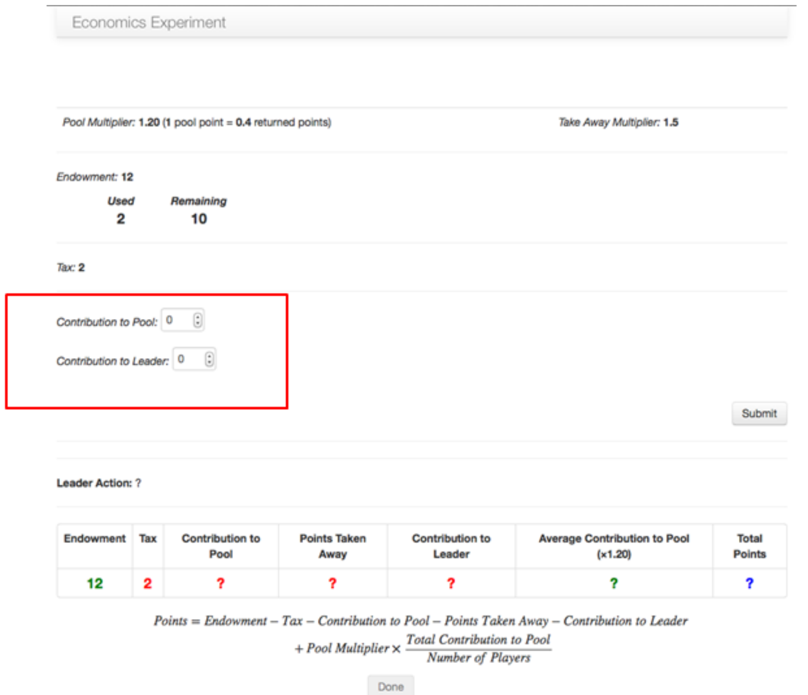
The first thing we'd like to draw your attention to is the pool multiplier. When group members allocate points to the group pool, the total allocated by the group is multipled by this number, the pool multiplier. In this case, it's 1.2. The new total is then distributed equally among everyone in the group, including the leader. This means that if every group member allocated all their points, all group members would increase their earnings by 20%. If no one allocates to the group pool, there's nothing to multiply by 1.2, and nothing to divide up. Another way to think about this is to consider how much each point in the pool is worth to each player. This is shown within parantheses to the right of the pool multiplier. In this case, with 3 players, each player gets 0.4 points for every point in the pool, no matter who contributes it. For example, if each player contributed 1 point pool, there would be 3 points in the pool and each player would get back 3 times 0.4 or 1.2 points.



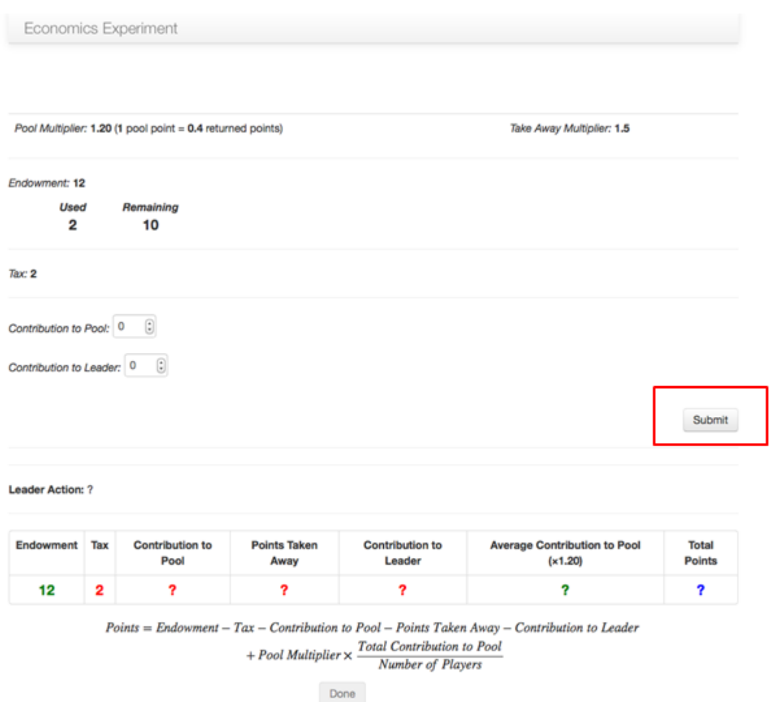
Below this, we see your initial endowment of 12 points. 2 points are taxed each period, which can be used by the leader to take points away from group members. This leaves you 10 points to allocate.



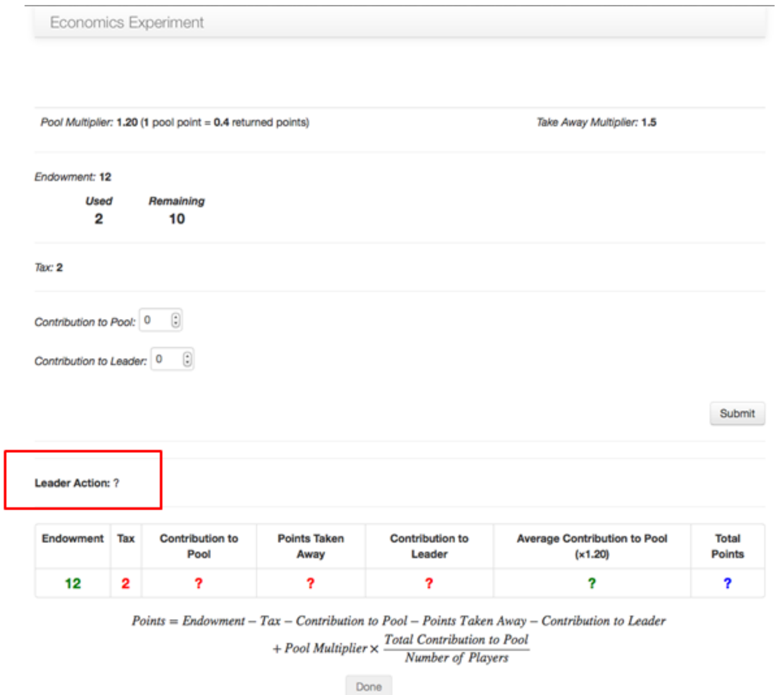
Here you see the Take Away multiplier. This is how much each tax point is multiplied by when the leader takes points away from a group member. For example, here, if the leader spent 1 tax point to take points away, 1.5 points would be taken from the targeted group member.



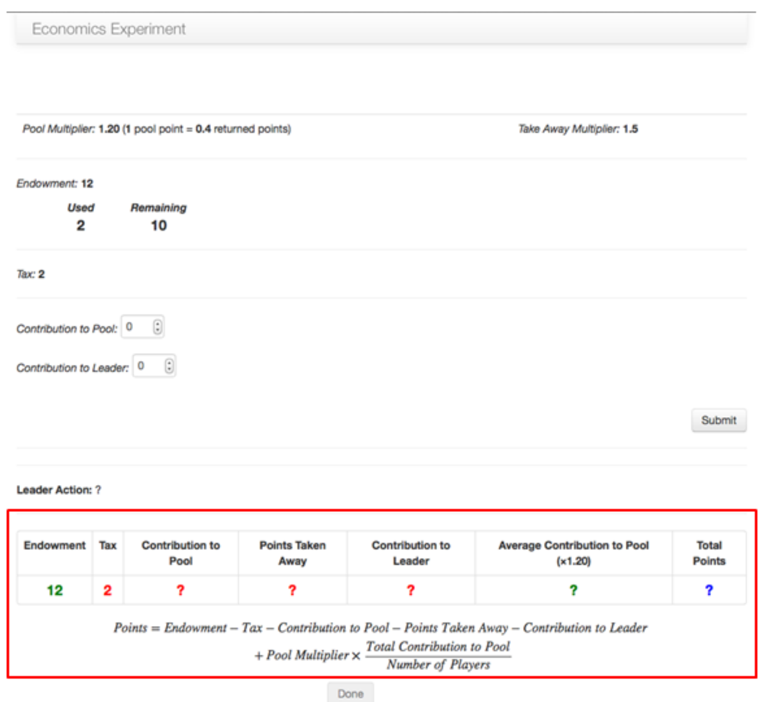
Here we see the boxes that are used to allocate your remaining 10 points to either the group pool or to the leader. Any points you do not allocate will be kept for yourself. When you hover your mouse over these boxes you will see an arrow to increase or decrease the amount in 1 point intervals or you may type in a value.



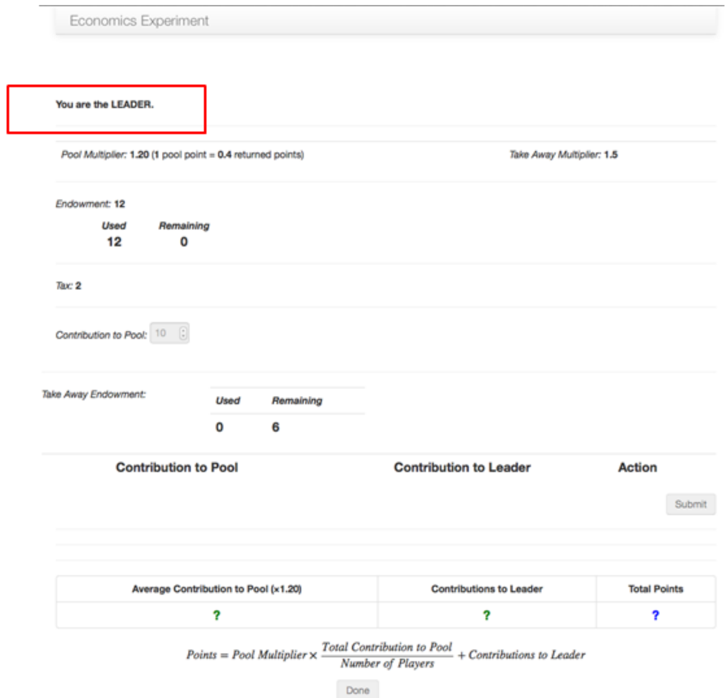
When you have made your decision, click “Submit”. After clicking “Submit”, you will not be able to change your decision.



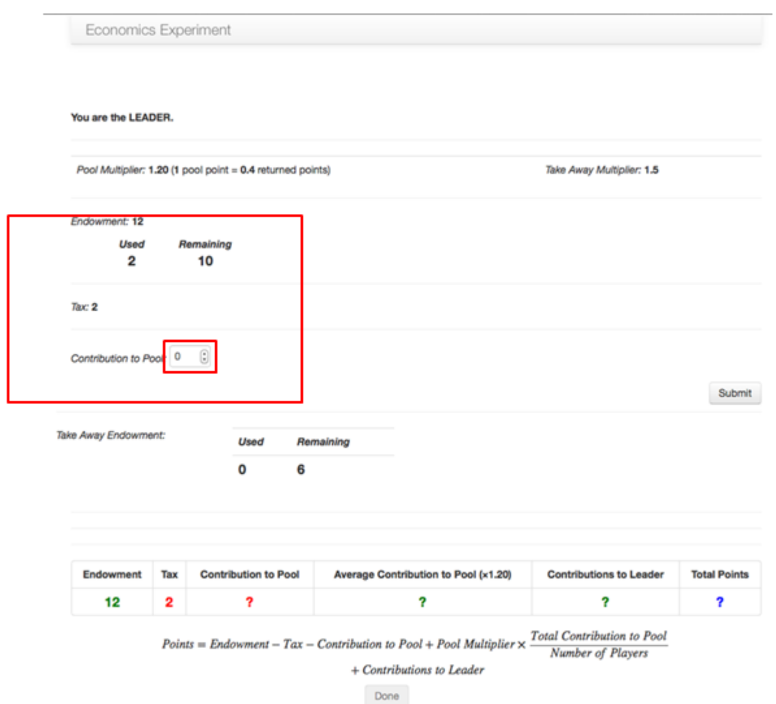
This section will show you the leader's action. The leader has 3 choices – to take points from you, accept your points, or do nothing. They may only choose one of these options.



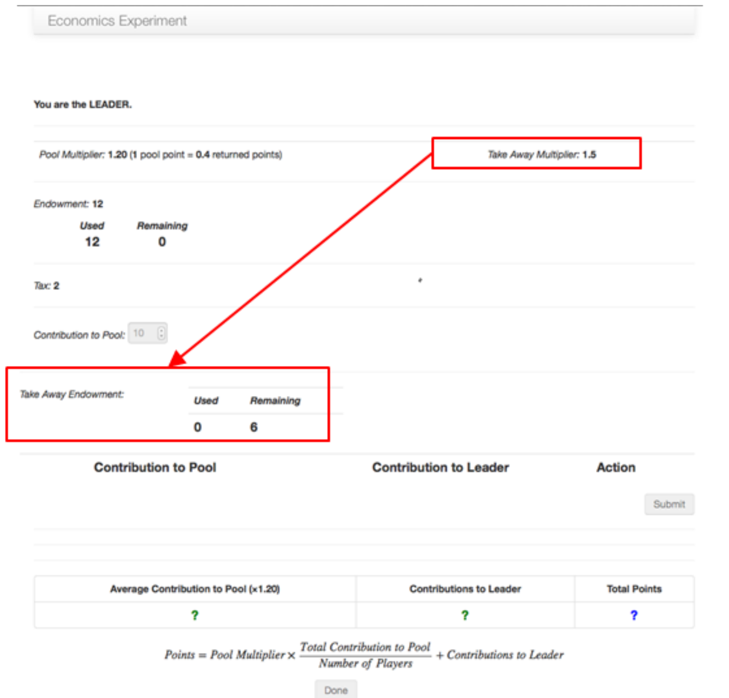
Here you see how your points for this period are calculated. Everything in green is added, everything in red is subtracted, and your total points for the period is shown in blue on the right. We will return to this after explaining the leader's screen.



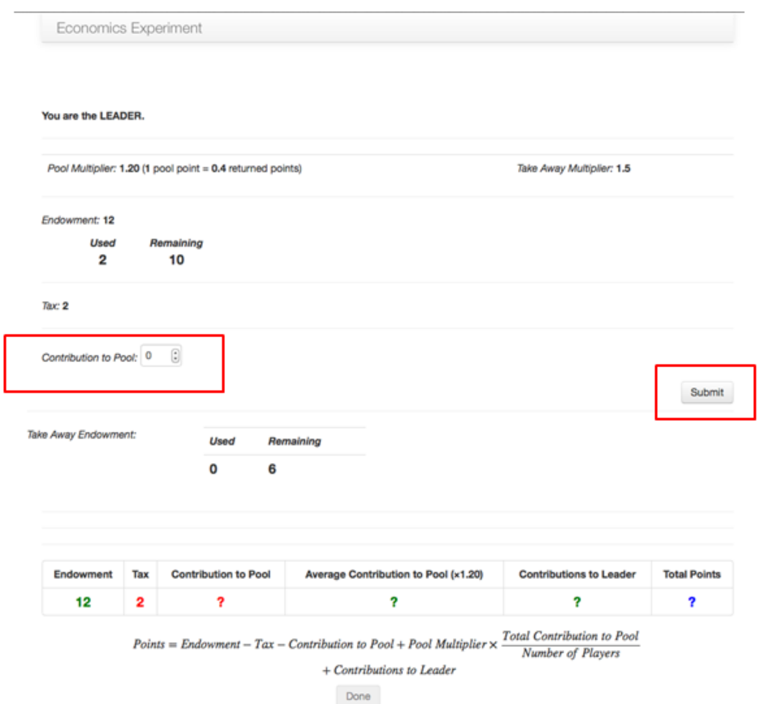
If you are the leader for a period, you will be identified at the top of the screen. Note that all group members, including the leader are anonymous to other group members. Your screen looks very similar to other group members, with a few differences.



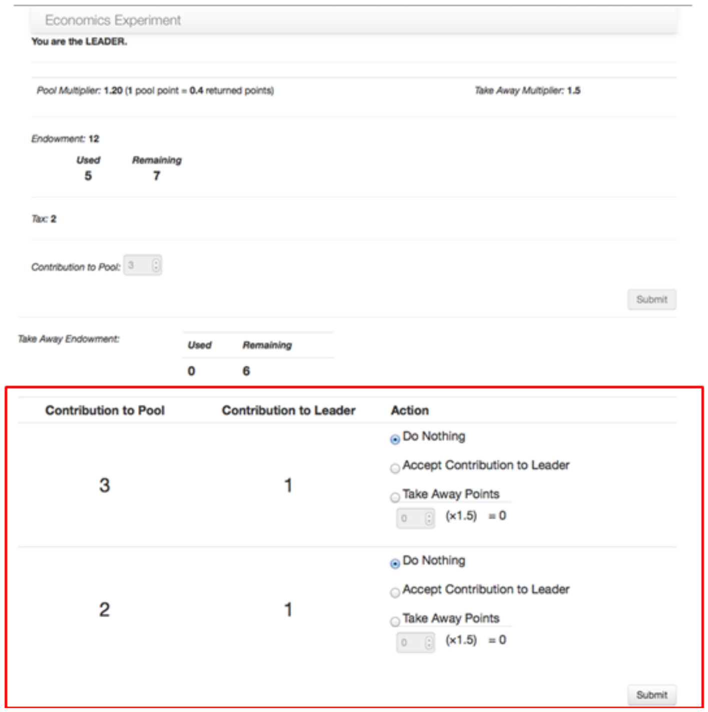
Like other group members you pay a tax and choose how much you wish to allocate to the group pool and how much to keep for yourself, but, unlike other group members you do not contribute to the leader, since you are the leader and you would be contributing to yourself.



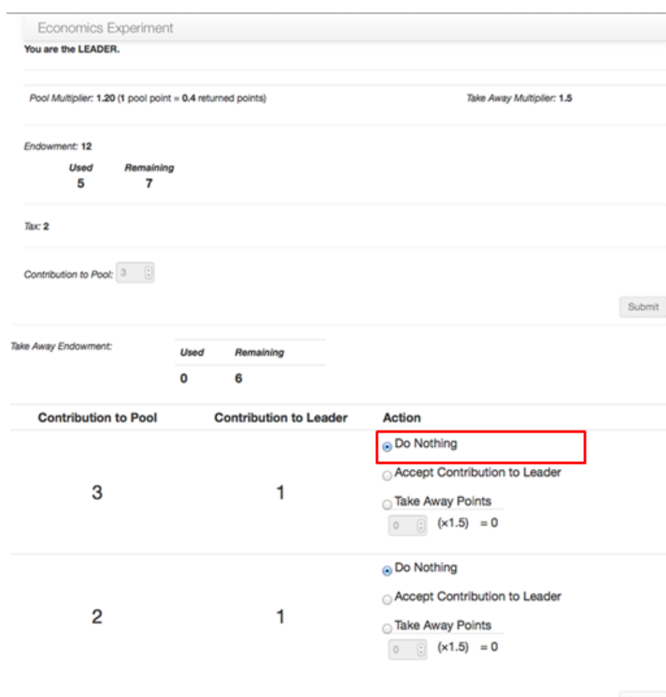
You also have access to tax points that you may use for taking away points from group members based on the Take Away multiplier.



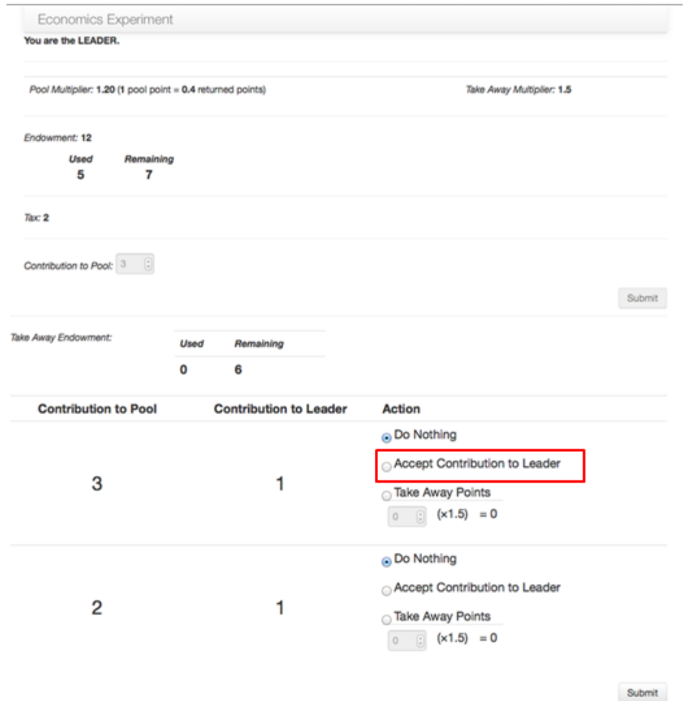
When you have made your decision about how many points to allocate to the group pool, click "Submit". After clicking "Submit", you will not be able to change your decision.



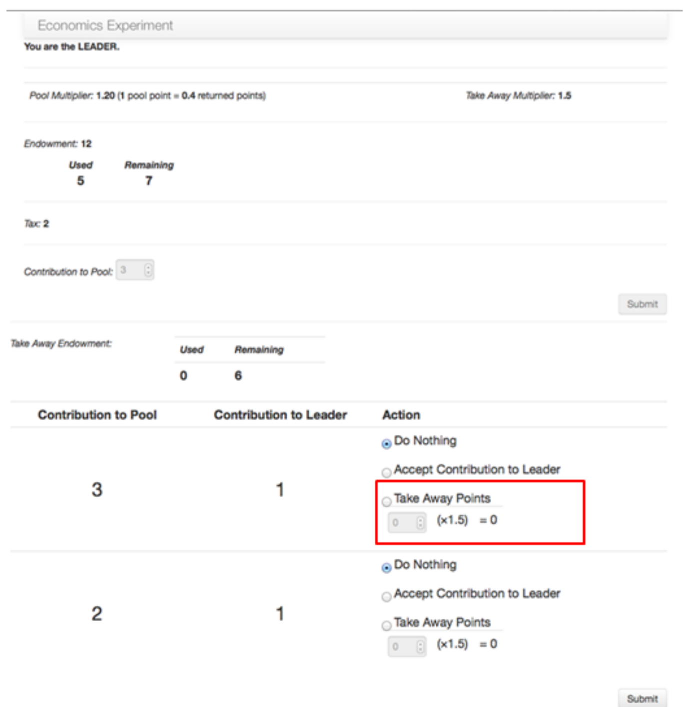
The actions of group members will appear below this after they make their decision. The Contribution column are the contributions to the group pool and the Payment column shows points given to you, the Leader.



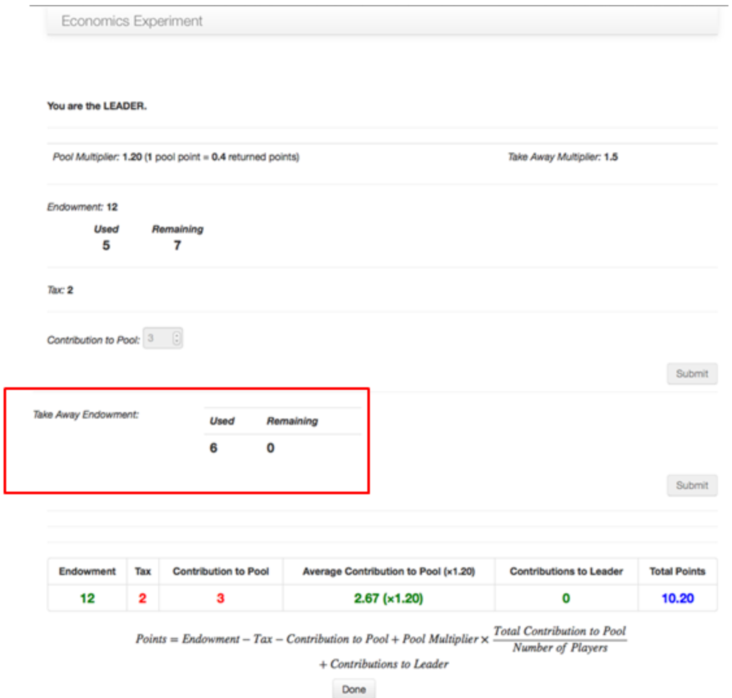
You may do nothing…



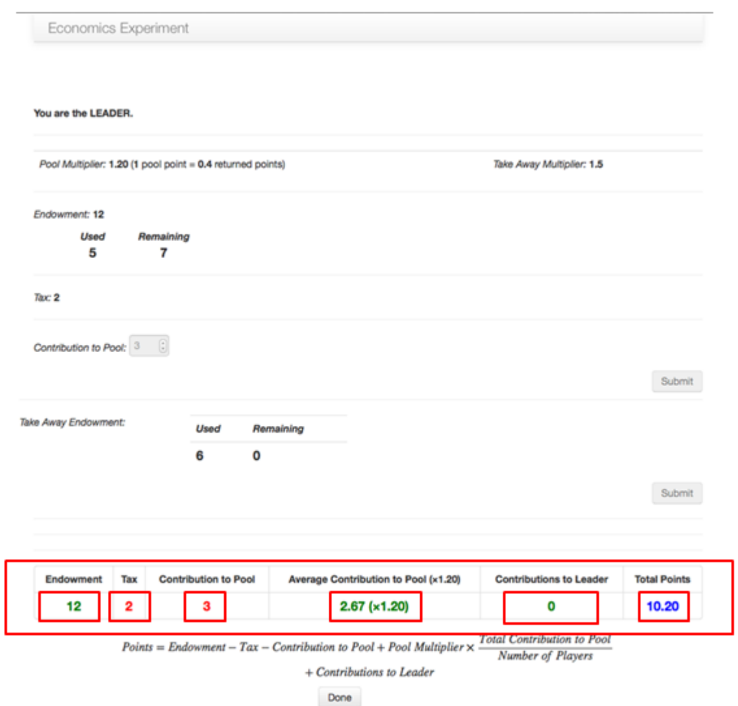
accept these points…



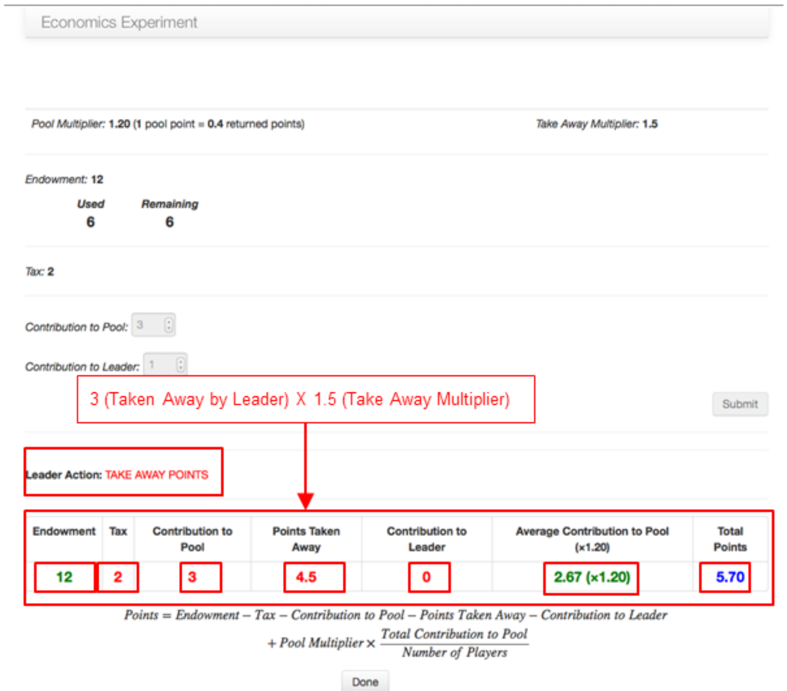
or take points away from any group member. If you take away points, you decide how much to take by allocating tax points to the group member. The total points taken away is calculated for you next to this box.



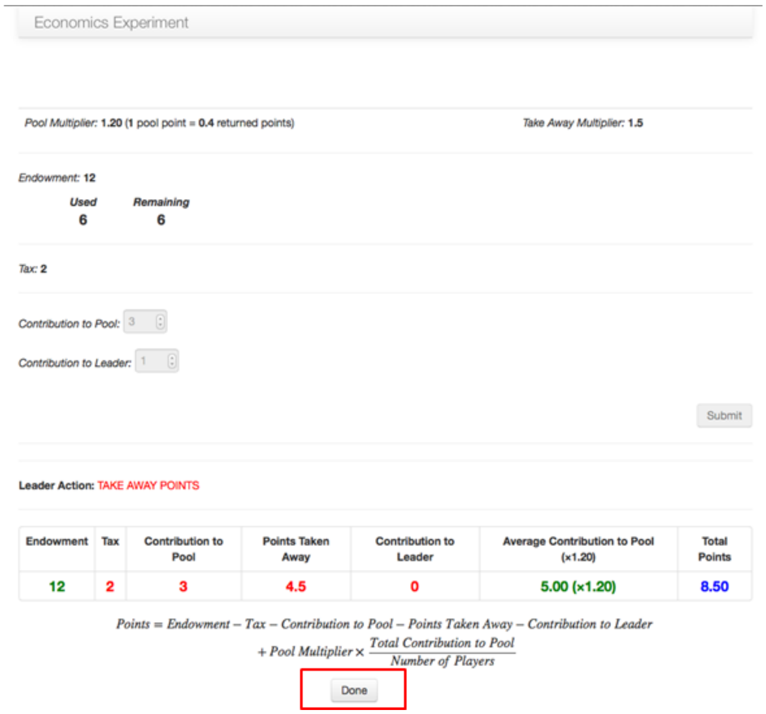
Any tax points not used for taking away points from group members will be returned to the experimenter. Similarly any points taken away from group members will also be returned to the experimenter.



As a leader your earnings for this period will be calculated from your initial endowment minus your taxes, minus your contribution, plus the average contribution to the pool – 2.67, multiplied by the pool multiplier, plus any points you accepted, giving you a total of 10.2 for this period.



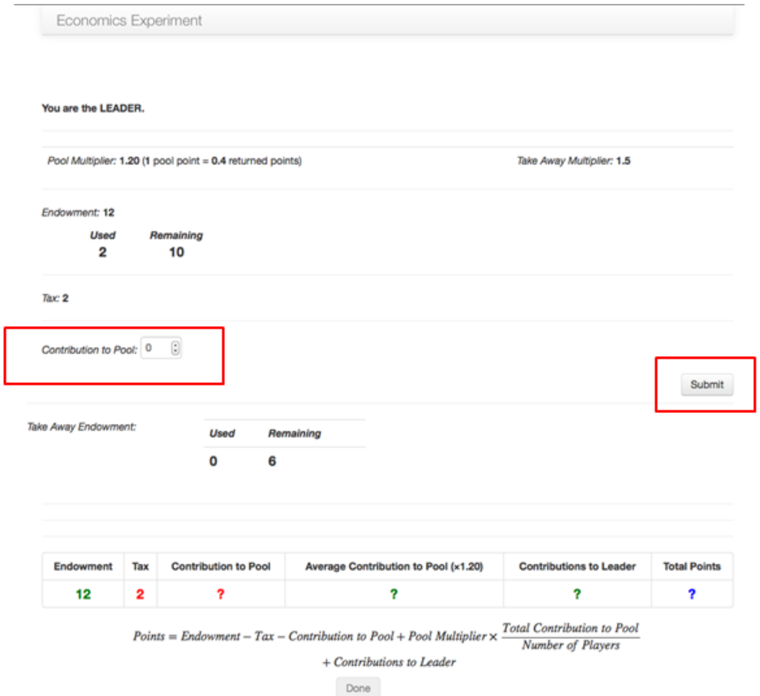
After the leader clicks submit, group members will be informed of the leader's action. The earnings of group members for this period will be calculated from their initial endowment minus their taxes, minus their contribution. Then either minus the points taken away by the leader OR minus the points given to the leader. Any points given to the leader and not accepted will be returned to the group member. In the example here, the leader chose to take points away from the group member. Finally, the group members share of the group pool will be added to this amount to calculate the total points for this period. Here the average contribution is shown – 2.67 points. This is multiplied by the pool multiplier and then added to give total points – 5.7 for this group member.



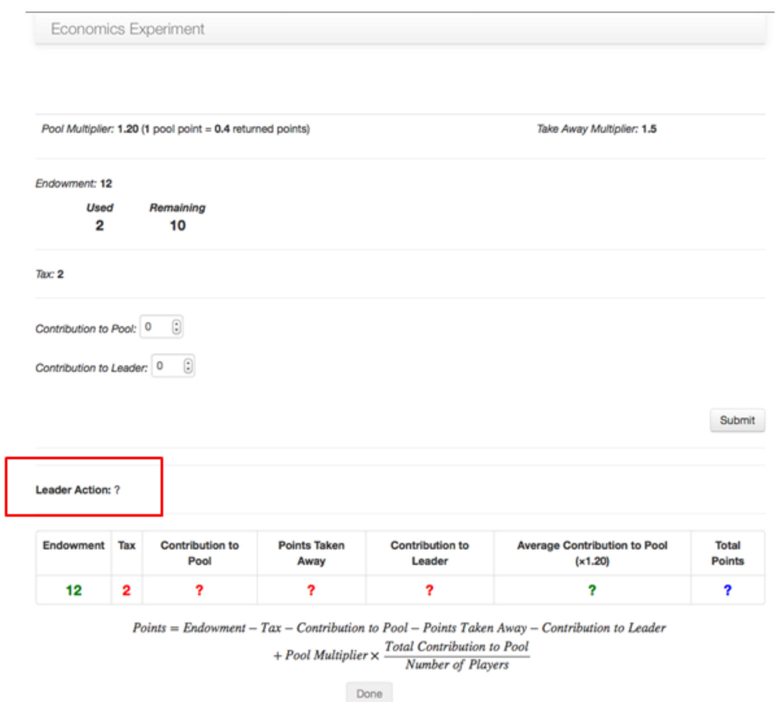
Please click "Done" at the bottom of the screen to move to the next period.

Please click "Continue" when you are ready to begin.

Subsequent Round



The difference between this round and previous rounds is that in this round, like other group members, the leader must choose how much to contribute to the group pool. They may also accept points from group members.



In addition, group members only see the leader's action towards them. They do not see the actions

of others group members or the leader's actions towards other group members. As in all rounds, all participants are completely anonymous.

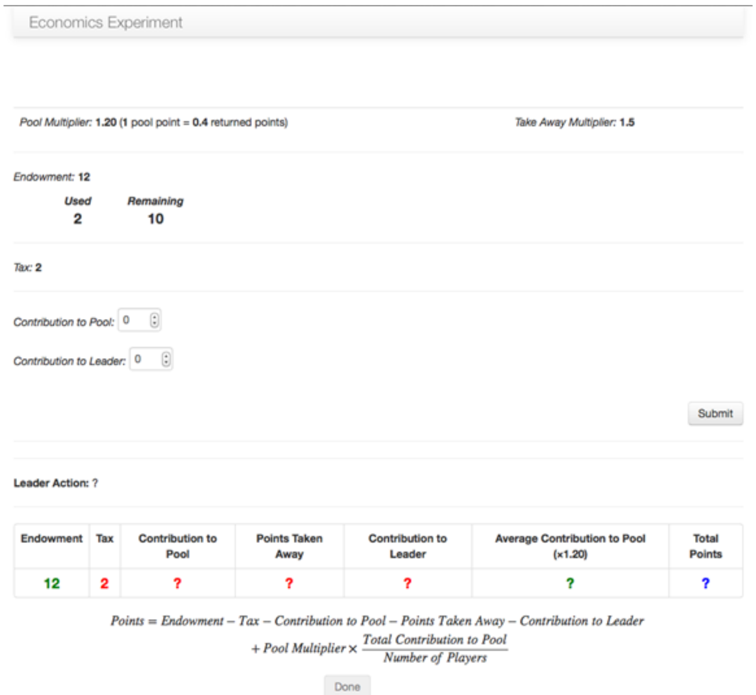
Please click "Continue" when you are ready to begin.

Quiz Questions

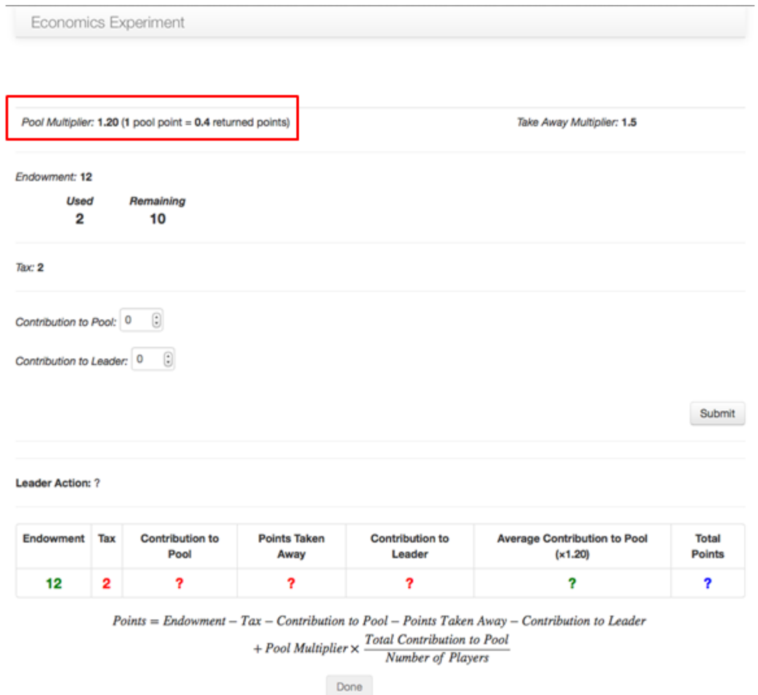
1. How many periods will you play?
    a. 10 periods
    b. More than 10 periods
    c. Less than 10 periods
    d. <u>An average of 10 periods, but possibly more and possibly less</u>
2. If everyone contributes the same amount to the group pool, will you:
    a. <u>Increase your points from your initial endowment</u>
    b. Decrease your points from your initial endowment
    c. Have the same points as your initial endowment
    d. Might increase your points or might decrease your points
3. How much does the leader contribute to the group pool?
    a. <u>However much they choose to allocate to the group pool</u>
    b. 10 points
    c. Depends on the points given to the leader
    d. Depends on the group member's contributions
4. Which of the following can the leader NOT do?
    a. Nothing
    b. Accept points from a group member
    c. Take points away from a group member
    d. <u>Accept points and take points away from the same group member</u>
5. If a leader does not accept points where do these points go?
    a. Goes to the leader anyway
    b. <u>Goes back to the group member</u>
    c. Goes to the experimenter
    d. Is added to the group pool
6. What actions of other group members can group members (not leader) see?
    a. <u>Leader's action towards them</u>
    b. Leader's contribution
    c. Leader's contribution and all actions
    d. All group member and leader actions

Bribery Game with Partial Transparency

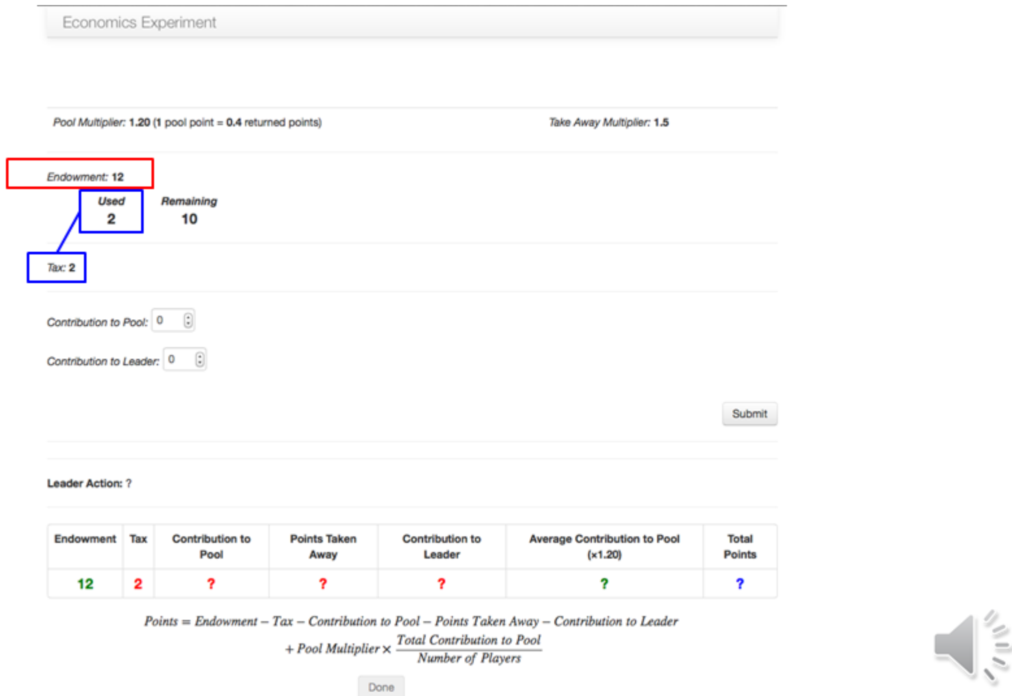
First Round



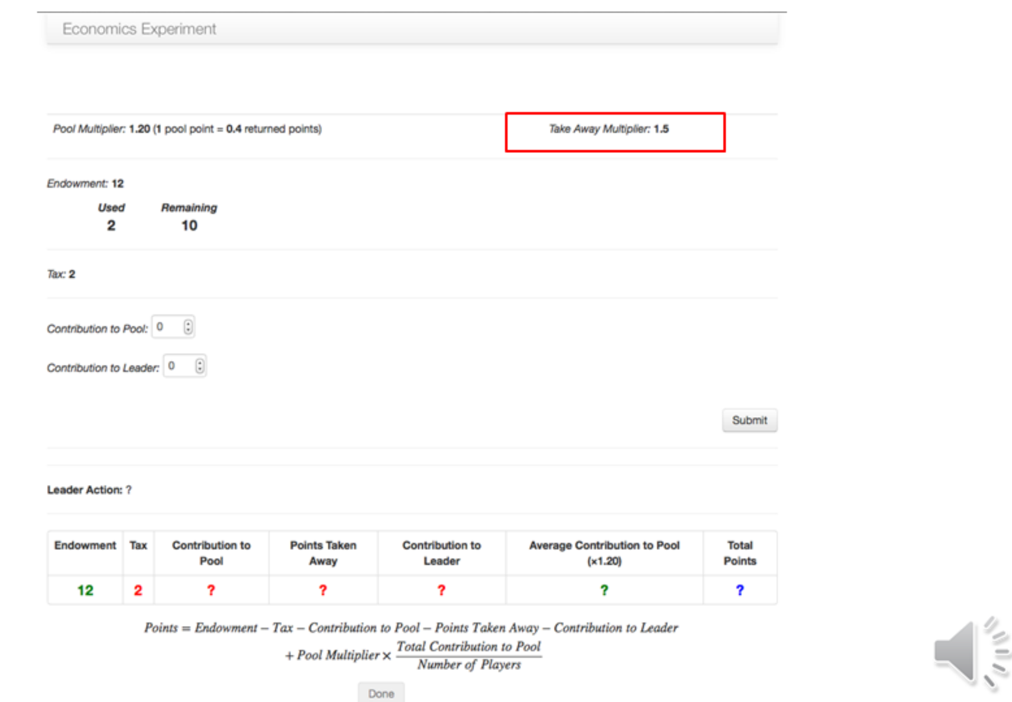
Welcome to the first round. The screen you see in front of you is the normal screen for play. In each period, one person from the group will be selected at random to play the role of a leader. The leader will have a slightly different screen. Please pay attention to all screens presented in these instructions as the instructions for other other rounds will only highlight the differences between that round and this one.



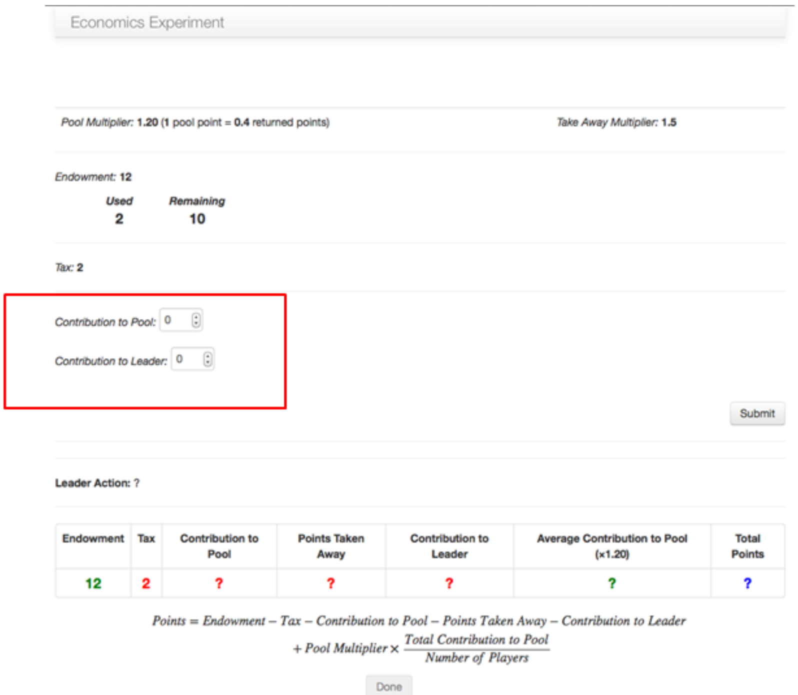
The first thing we'd like to draw your attention to is the pool multiplier. When group members allocate points to the group pool, the total allocated by the group is multipled by this number, the pool multiplier. In this case, it's 1.2. The new total is then distributed equally among everyone in the group, including the leader. This means that if every group member allocated all their points, all group members would increase their earnings by 20%. If no one allocates to the group pool, there's nothing to multiply by 1.2, and nothing to divide up. Another way to think about this is to consider how much each point in the pool is worth to each player. This is shown within parantheses to the right of the pool multiplier. In this case, with 3 players, each player gets 0.4 points for every point in the pool, no matter who contributes it. For example, if each player contributed 1 point pool, there would be 3 points in the pool and each player would get back 3 times 0.4 or 1.2 points.



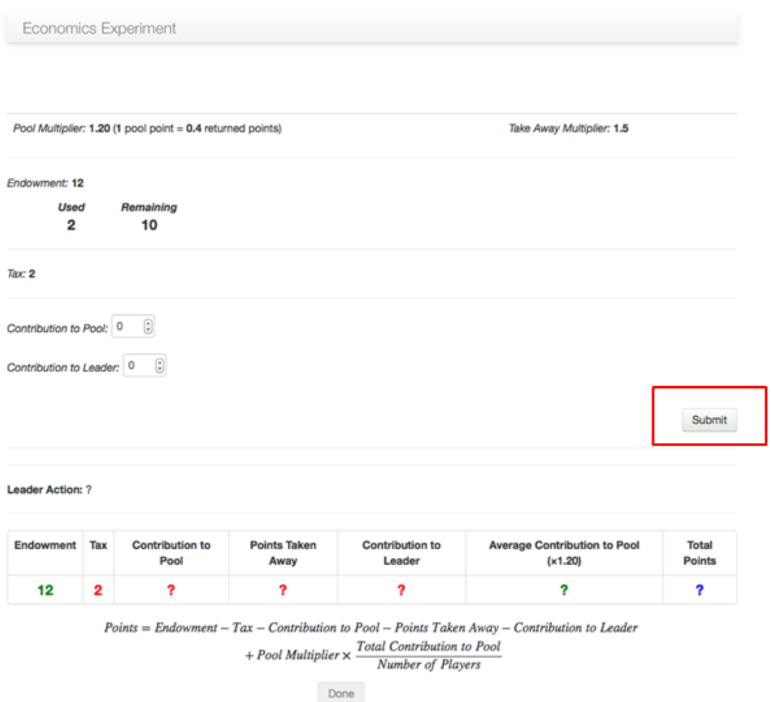
Below this, we see your initial endowment of 12 points. 2 points are taxed each period, which can be used by the leader to take points away from group members. This leaves you 10 points to allocate.



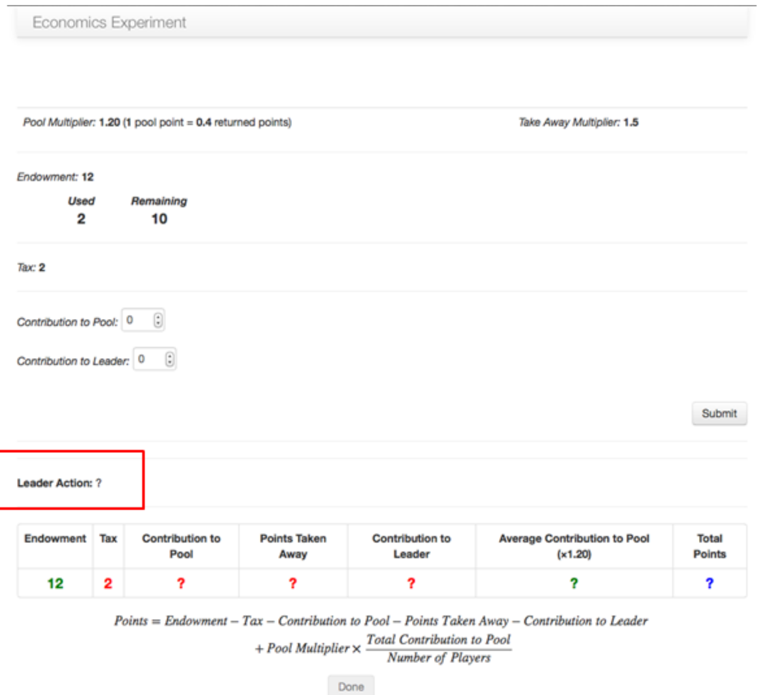
Here you see the Take Away multiplier. This is how much each tax point is multiplied by when the leader takes points away from a group member. For example, here, if the leader spent 1 tax point to take points away, 1.5 points would be taken from the targeted group member.



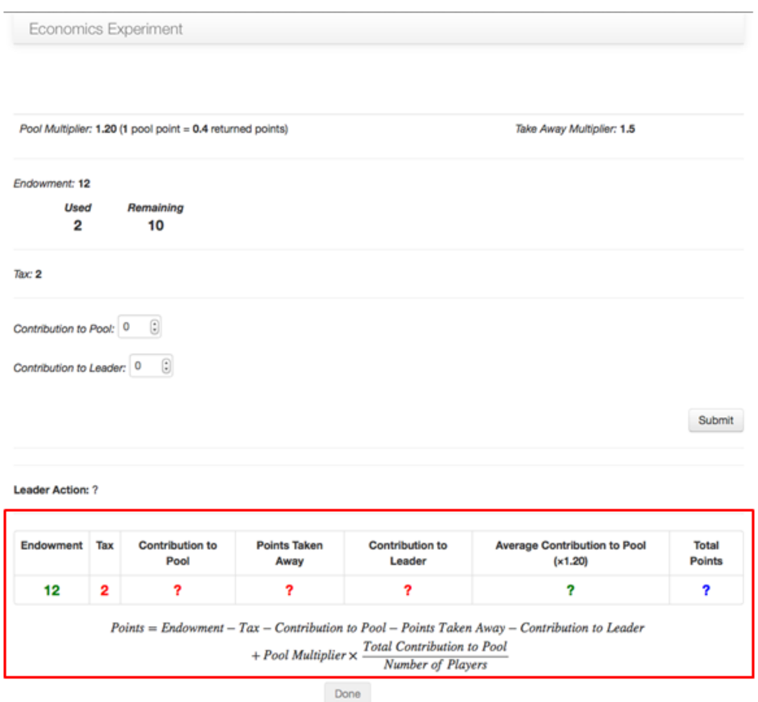
Here we see the boxes that are used to allocate your remaining 10 points to either the group pool or to the leader. Any points you do not allocate will be kept for yourself. When you hover your mouse over these boxes you will see an arrow to increase or decrease the amount in 1 point intervals or you may type in a value.



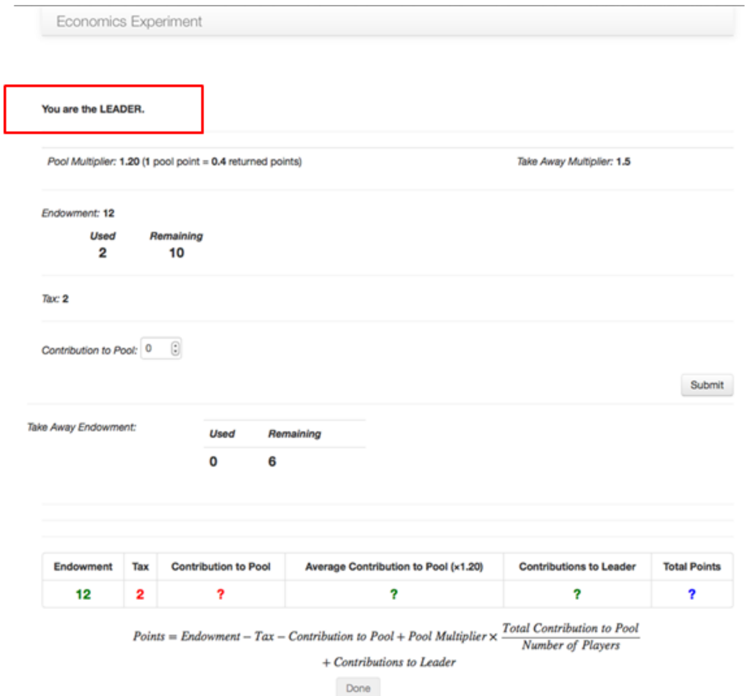
When you have made your decision, click "Submit". After clicking "Submit", you will not be able to change your decision.



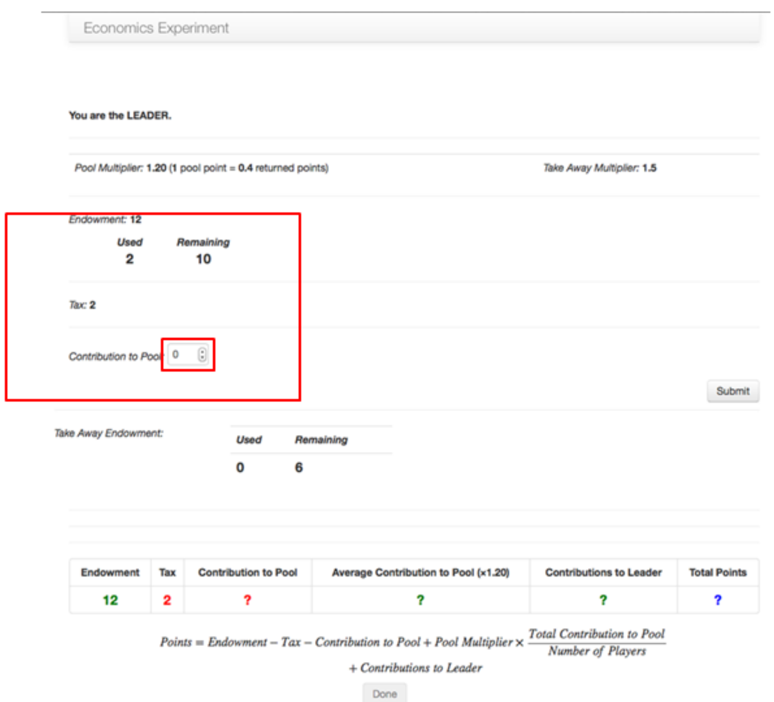
This section will show you the leader's action. The leader has 3 choices – to take points from you, accept your points, or do nothing. They may only choose one of these options.



Here you see how your points for this period are calculated. Everything in green is added, everything in red is subtracted, and your total points for the period is shown in blue on the right. We will return to this after explaining the leader's screen.

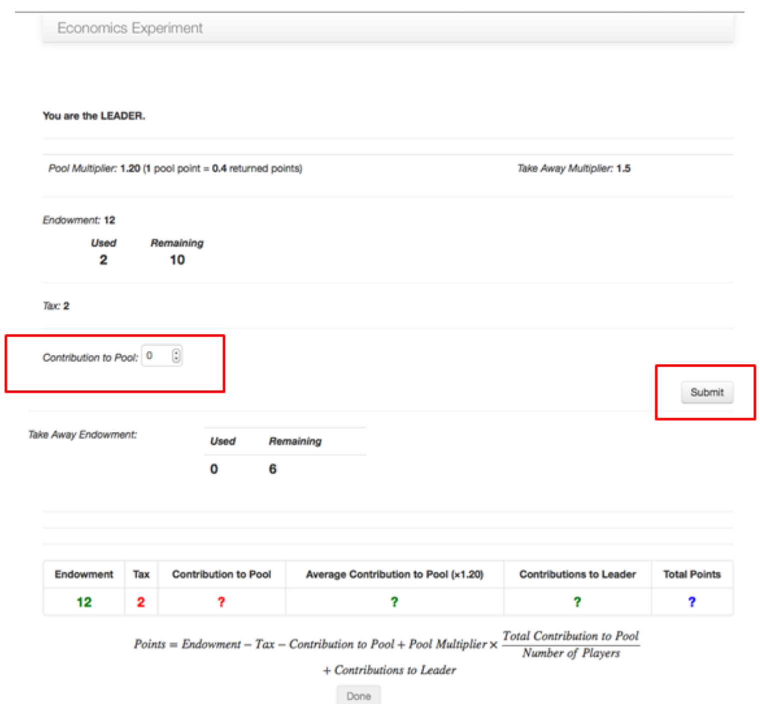


If you are the leader for a period, you will be identified at the top of the screen. Note that all group members, including the leader are anonymous to other group members. Your screen looks very similar to other group members, with a few differences.

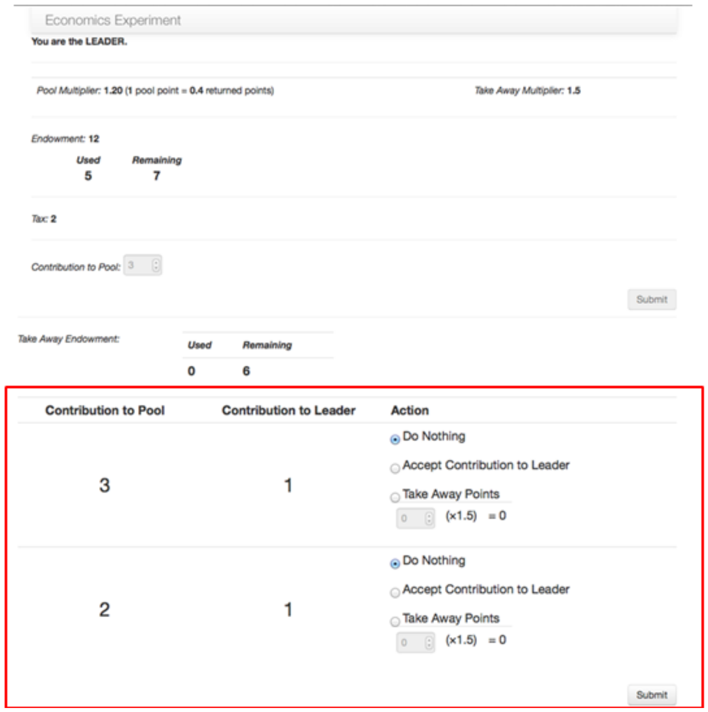


Like other group members you pay a tax and choose how much you wish to allocate to the group pool and how much to keep for yourself, but, unlike other group members you do not contribute to the leader, since you are the leader and you would be contributing to yourself.

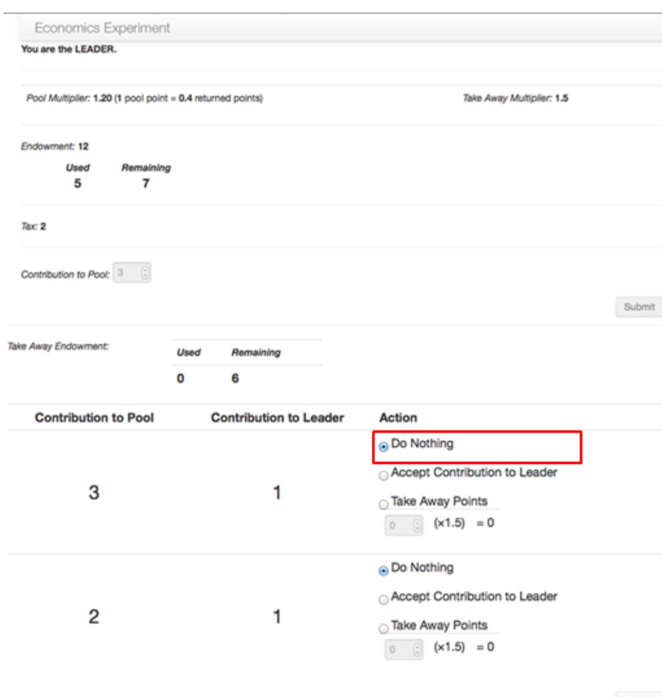
You also have access to tax points that you may use for taking away points from group members based on the Take Away multiplier.



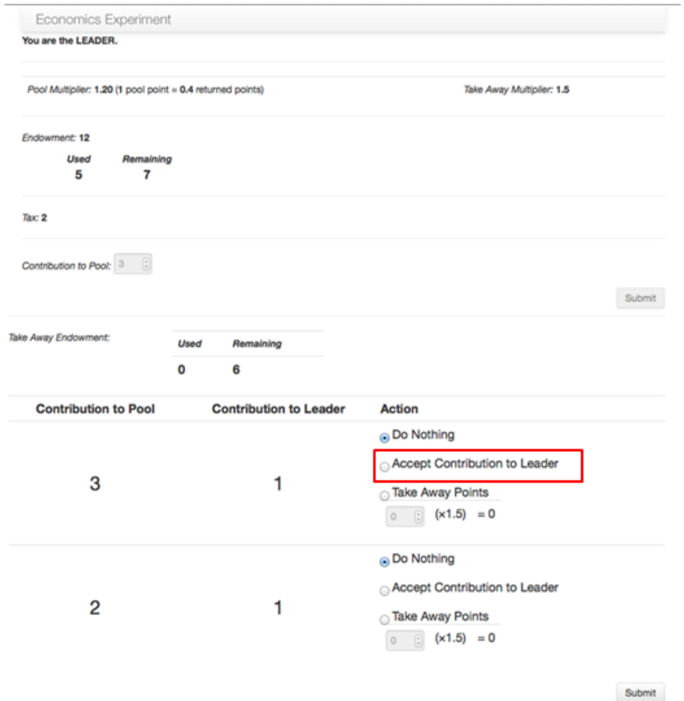
When you have made your decision about how many points to allocate to the group pool, click "Submit". After clicking "Submit", you will not be able to change your decision.



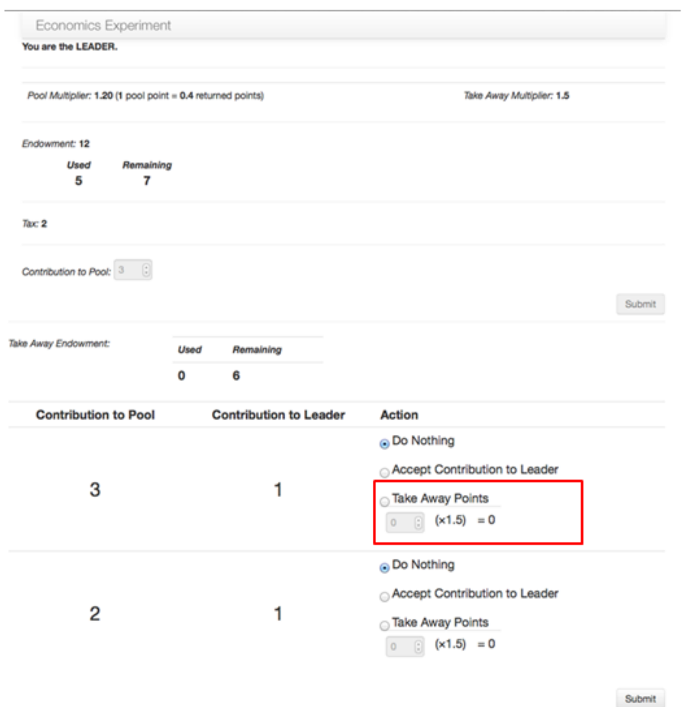
The actions of group members will appear below this after they make their decision. The Contribution column are the contributions to the group pool and the Payment column shows points given to you, the Leader.



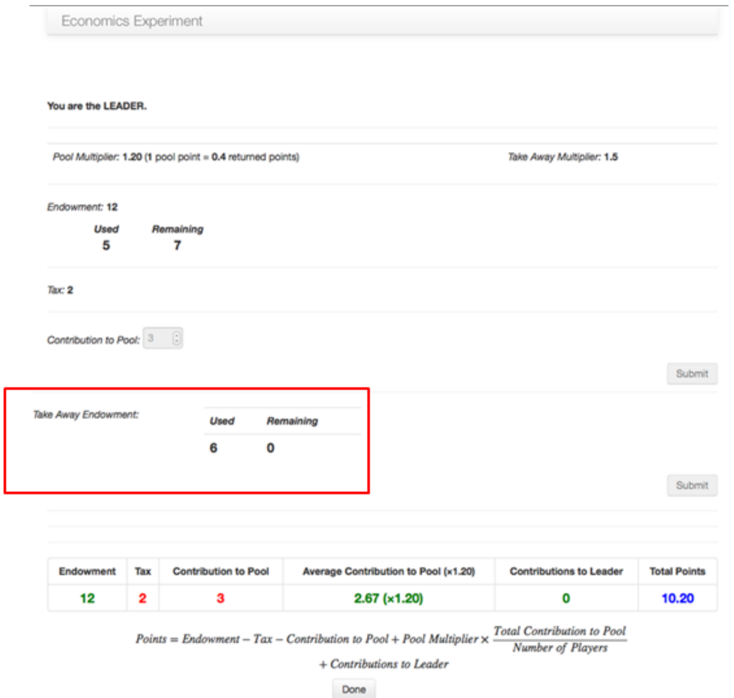
You may do nothing…



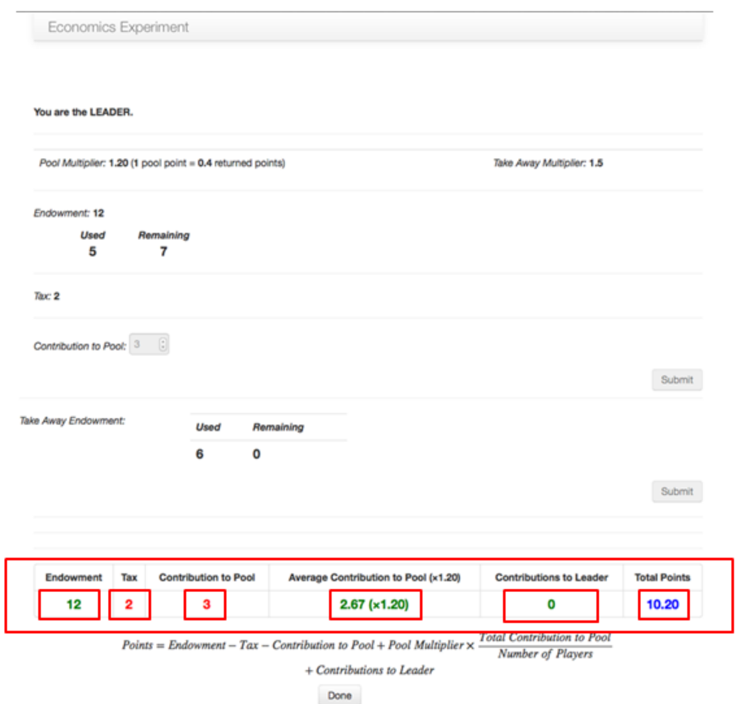
accept these points…



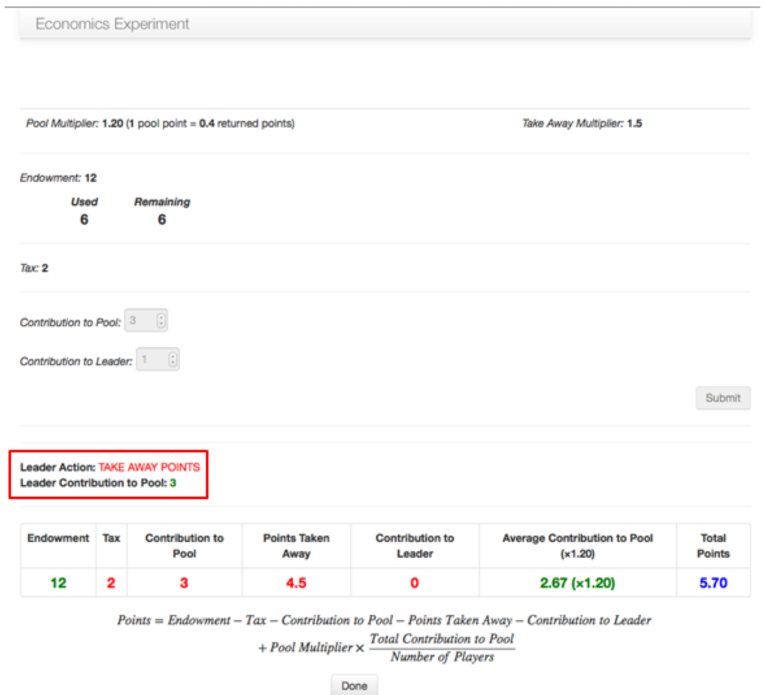
or take points away from any group member. If you take away points, you decide how much to take by allocating tax points to the group member. The total points taken away is calculated for you next to this box.



Any tax points not used for taking away points from group members will be returned to the experimenter. Similarly any points taken away from group members will also be returned to the experimenter.



As a leader your earnings for this period will be calculated from your initial endowment minus your taxes, minus your contribution, plus the average contribution to the pool – 2.67, multiplied by the pool multiplier, plus any points you accepted, giving you a total of 10.2 for this period.

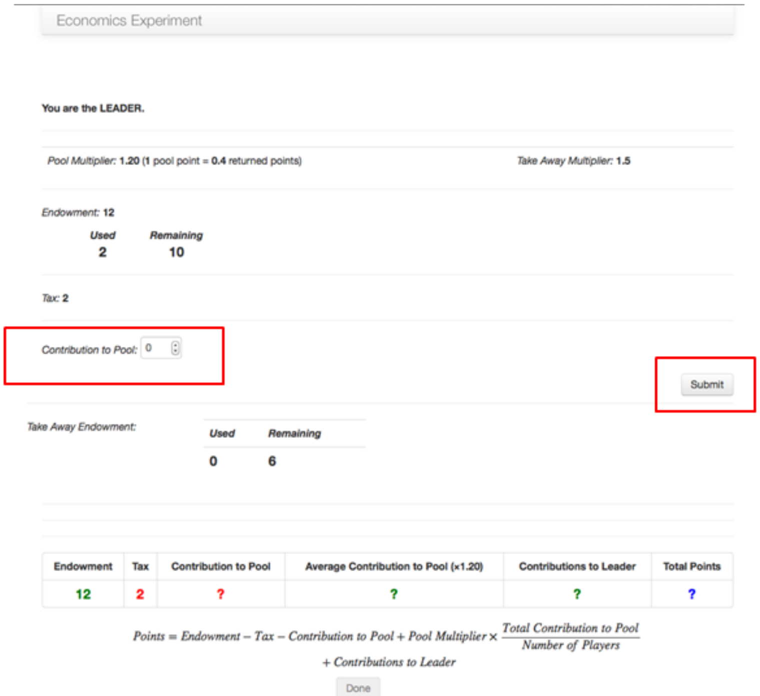


After the leader clicks submit, group members will be informed of the leader's action and the leader's contribution to the pool. The earnings of group members for this period will be calculated from their initial endowment minus their taxes, minus their contribution. Then either minus the points taken away by the leader OR minus the points given to the leader. Any points given to the leader and not accepted will be returned to the group member. In the example here, the leader chose to take points away from the group member. Finally, the group members share of the group pool will be added to this amount to calculate the total points for this period. Here the average contribution is shown – 2.67 points. This is multiplied by the pool multiplier and then added to give total points – 5.7 for this group member.

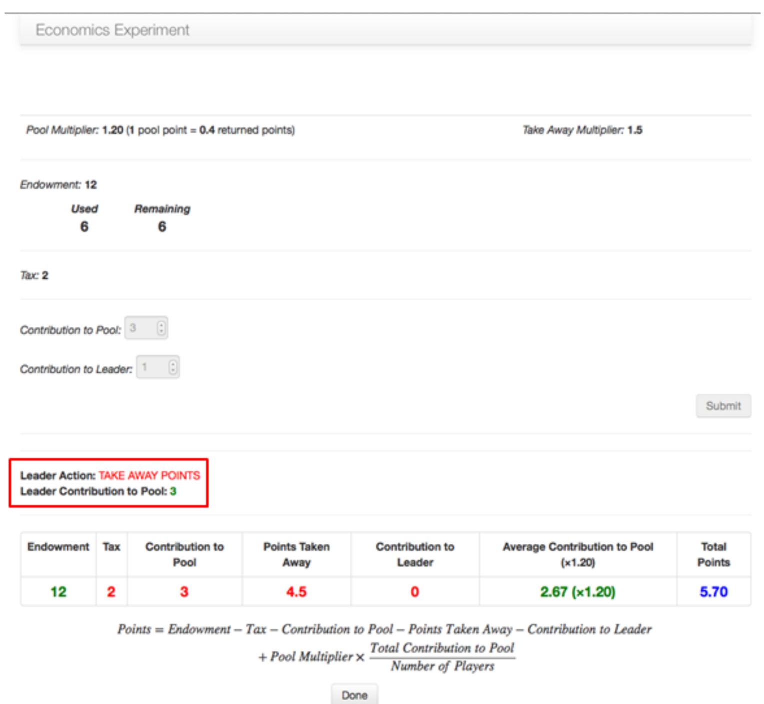
Please click "Done" at the bottom of the screen to move to the next period.

Please click "Continue" when you are ready to begin.

Subsequent Round



The difference between this round and previous rounds is that in this round, like other group members, the leader must choose how much to contribute to the group pool. They may also accept points from group members.



In addition, group members not only see the leader's action towards them, but also the leader's contribution to the group pool. They do not see the leader's actions towards other group members. As in all rounds, all participants are completely anonymous.

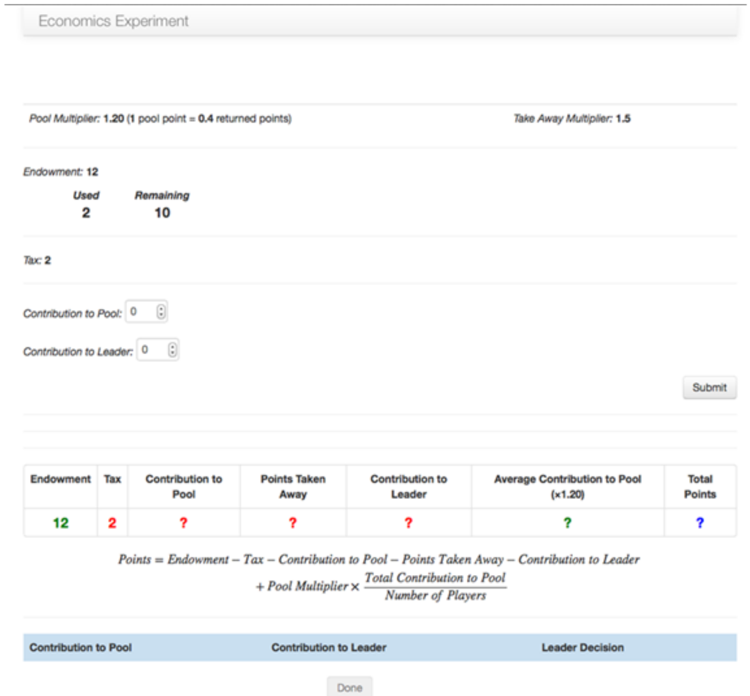
Please click "Continue" when you are ready to begin.

Quiz Questions

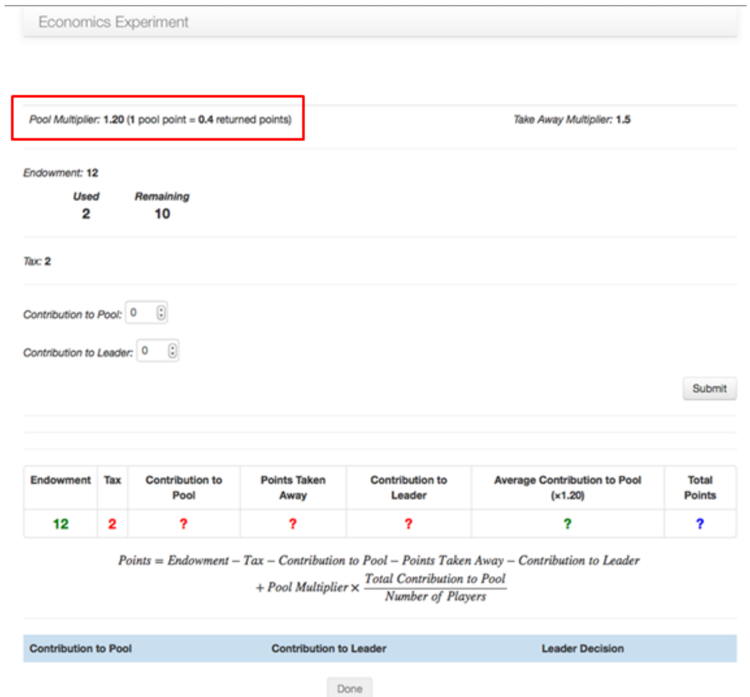
1. How many periods will you play?
    a. 10 periods
    b. More than 10 periods
    c. Less than 10 periods
    d. <u>An average of 10 periods, but possibly more and possibly less</u>
2. If everyone contributes the same amount to the group pool, will you:
    a. <u>Increase your points from your initial endowment</u>
    b. Decrease your points from your initial endowment
    c. Have the same points as your initial endowment
    d. Might increase your points or might decrease your points
3. How much does the leader contribute to the group pool?
    a. <u>However much they choose to allocate to the group pool</u>
    b. 10 points
    c. Depends on the points given to the leader
    d. Depends on the group member's contributions
4. Which of the following can the leader NOT do?
    a. Nothing
    b. Accept points from a group member
    c. Take points away from a group member
    d. <u>Accept points and take points away from the same group member</u>
5. If a leader does not accept points where do these points go?
    a. Goes to the leader anyway
    b. <u>Goes back to the group member</u>
    c. Goes to the experimenter
    d. Is added to the group pool
6. What actions of other group members can group members see?
    a. Leader's action towards them
    b. Leader's contribution
    c. <u>Leader's contribution and all actions</u>
    d. All group member and leader actions

Bribery Game with Full Transparency

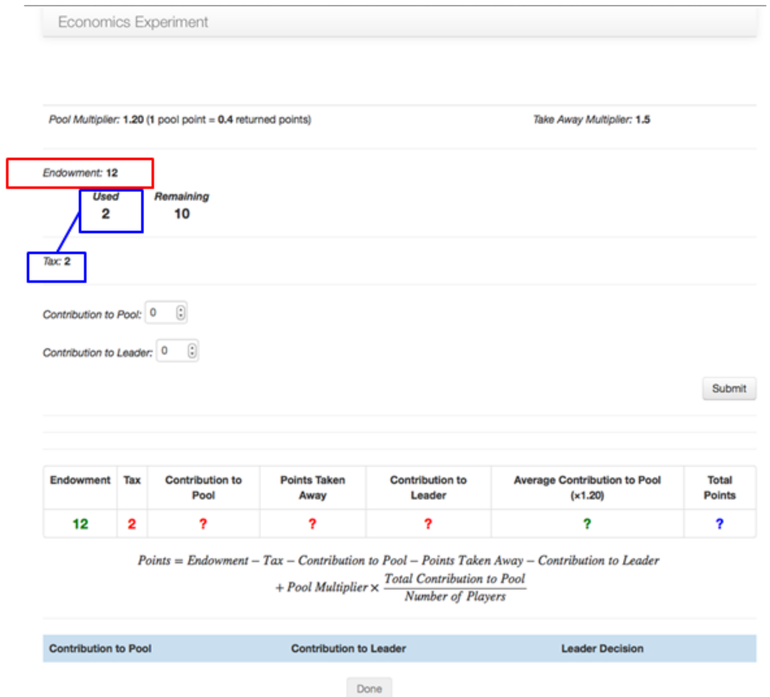
First Round



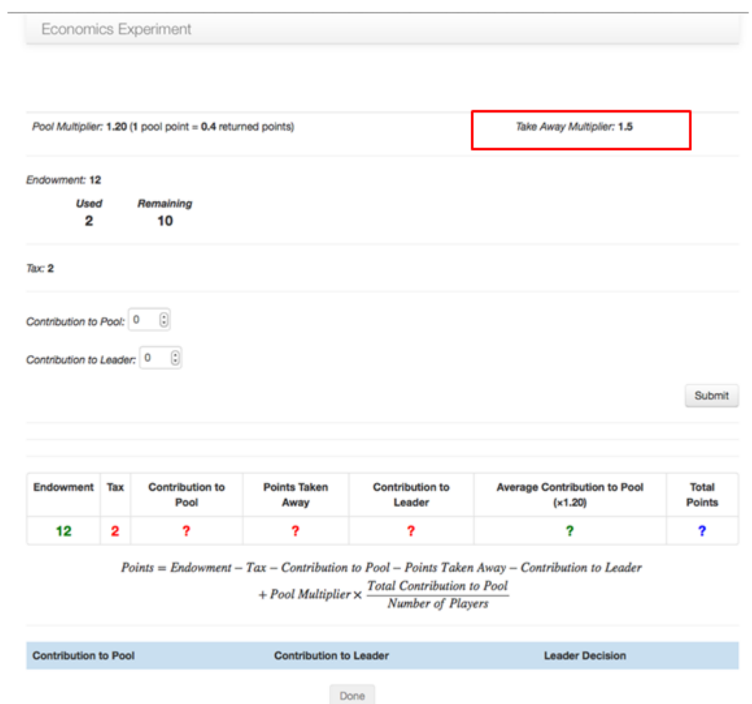
Welcome to the first round. The screen you see in front of you is the normal screen for play. In each period, one person from the group will be selected at random to play the role of a leader. The leader will have a slightly different screen. Please pay attention to all screens presented in these instructions as the instructions for other other rounds will only highlight the differences between that round and this one.



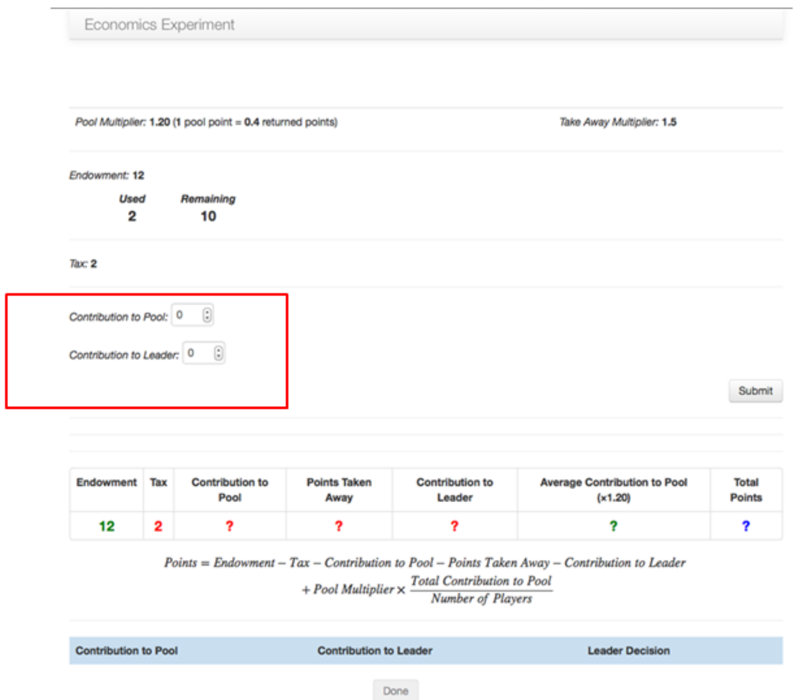
The first thing we'd like to draw your attention to is the pool multiplier. When group members allocate points to the group pool, the total allocated by the group is multipled by this number, the pool multiplier. In this case, it's 1.2. The new total is then distributed equally among everyone in the group, including the leader. This means that if every group member allocated all their points, all group members would increase their earnings by 20%. If no one allocates to the group pool, there's nothing to multiply by 1.2, and nothing to divide up. Another way to think about this is to consider how much each point in the pool is worth to each player. This is shown within parantheses to the right of the pool multiplier. In this case, with 3 players, each player gets 0.4 points for every point in the pool, no matter who contributes it. For example, if each player contributed 1 point pool, there would be 3 points in the pool and each player would get back 3 times 0.4 or 1.2 points.



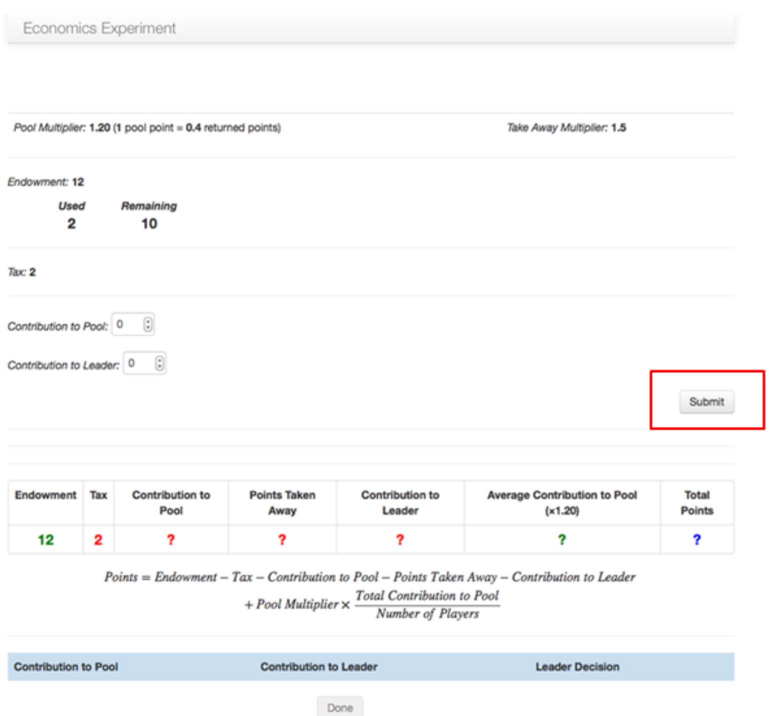
Below this, we see your initial endowment of 12 points. 2 points are taxed each period, which can be used by the leader to take points away from group members. This leaves you 10 points to allocate.



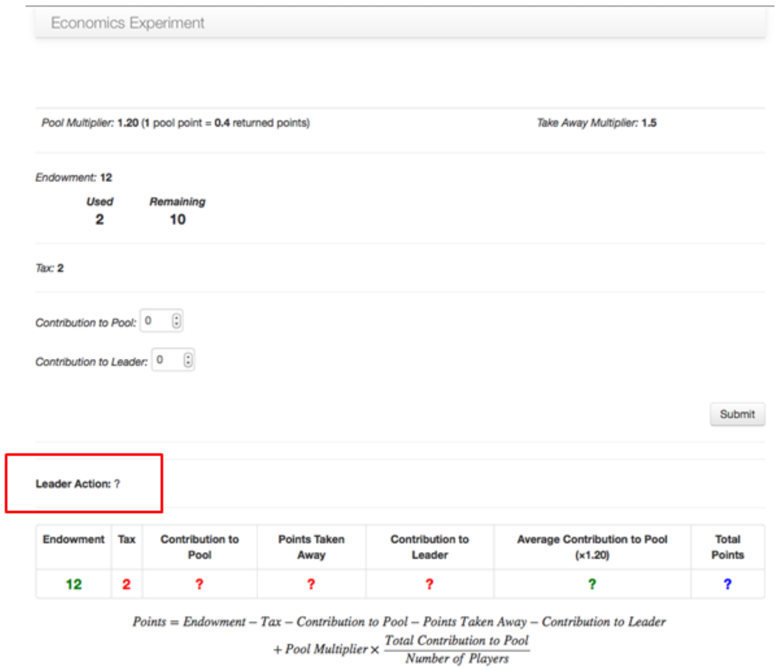
Here you see the Take Away multiplier. This is how much each tax point is multiplied by when the leader takes points away from a group member. For example, here, if the leader spent 1 tax point to take points away, 1.5 points would be taken from the targeted group member.



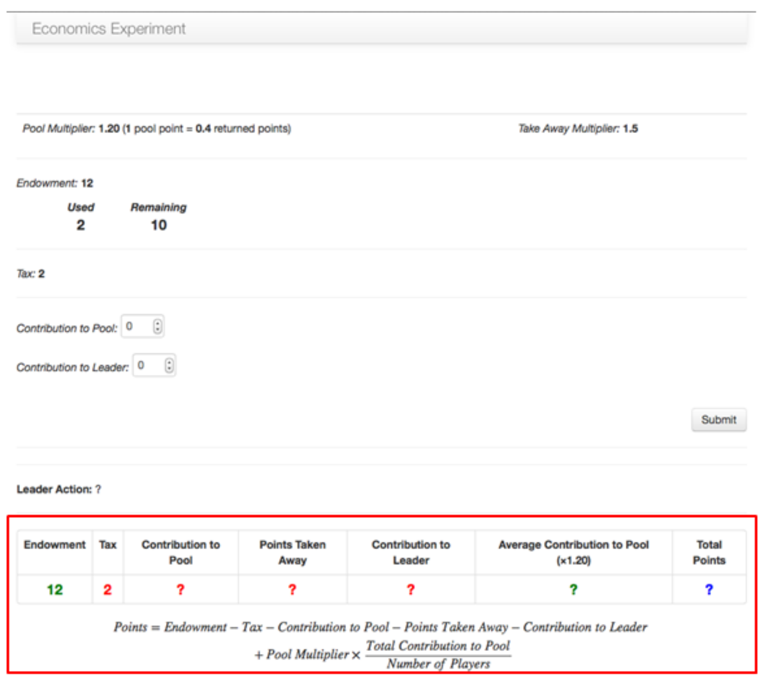
Here we see the boxes that are used to allocate your remaining 10 points to either the group pool or to the leader. Any points you do not allocate will be kept for yourself. When you hover your mouse over these boxes you will see an arrow to increase or decrease the amount in 1 point intervals or you may type in a value.



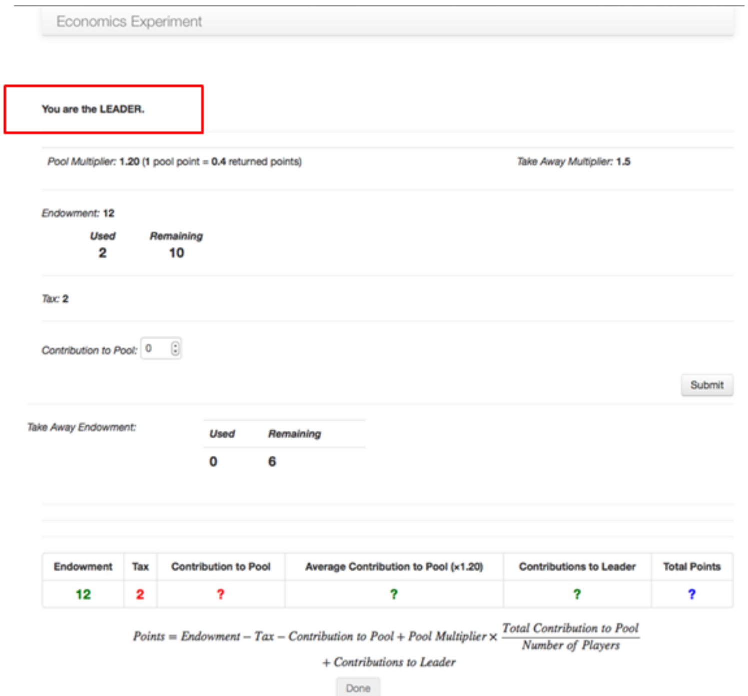
When you have made your decision, click "Submit". After clicking "Submit", you will not be able to change your decision.



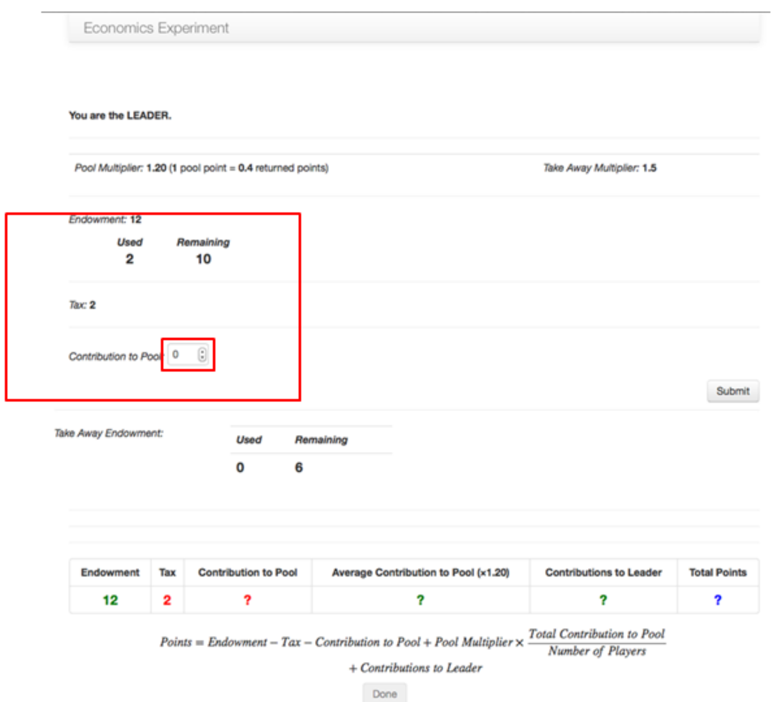
This section will show you the leader's action. The leader has 3 choices – to take points from you, accept your points, or do nothing. They may only choose one of these options.



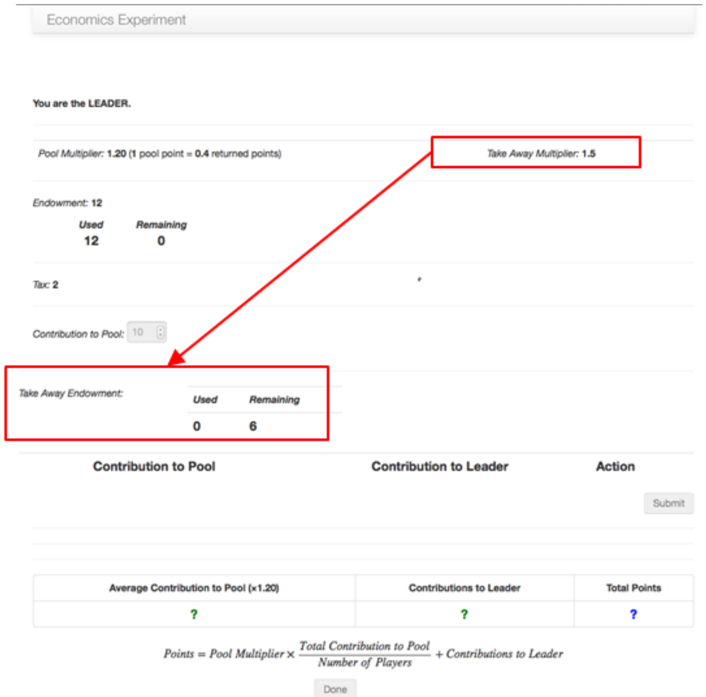
Here you see how your points for this period are calculated. Everything in green is added, everything in red is subtracted, and your total points for the period is shown in blue on the right. We will return to this after explaining the leader's screen.



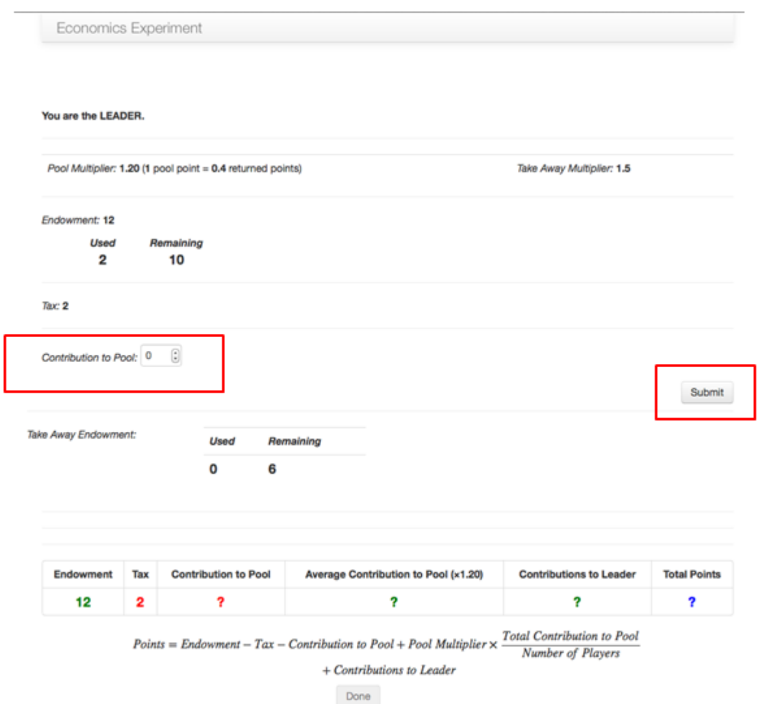
If you are the leader for a period, you will be identified at the top of the screen. Note that all group members, including the leader are anonymous to other group members. Your screen looks very similar to other group members, with a few differences.



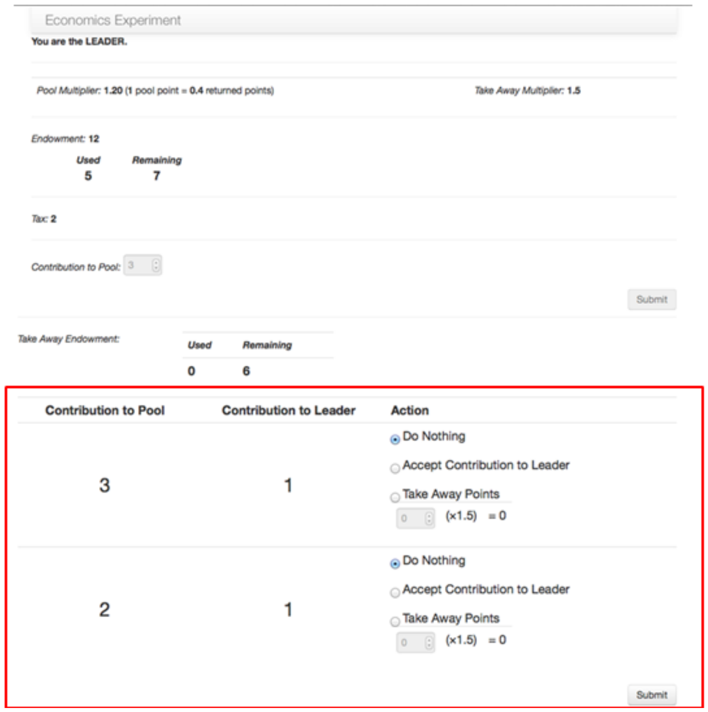
Like other group members you pay a tax and choose how much you wish to allocate to the group pool and how much to keep for yourself, but, unlike other group members you do not contribute to the leader, since you are the leader and you would be contributing to yourself.



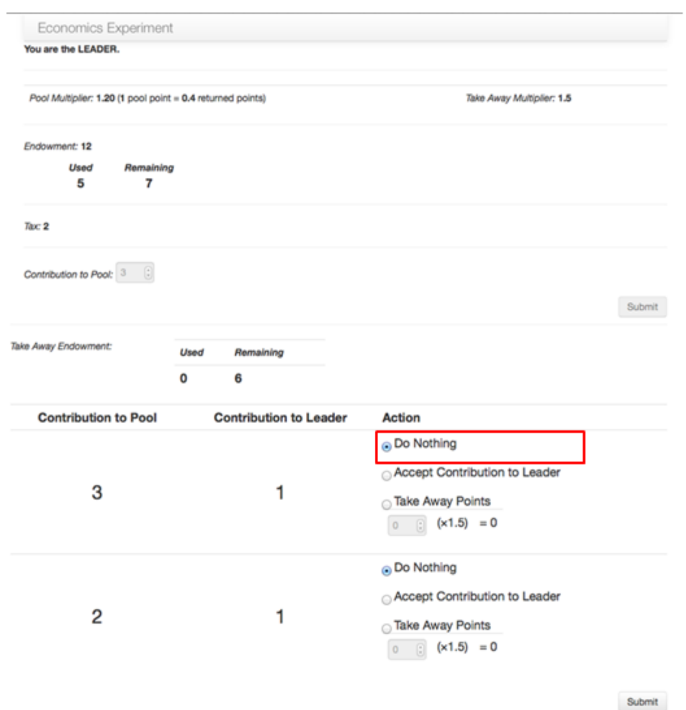
You also have access to tax points that you may use for taking away points from group members based on the Take Away multiplier.



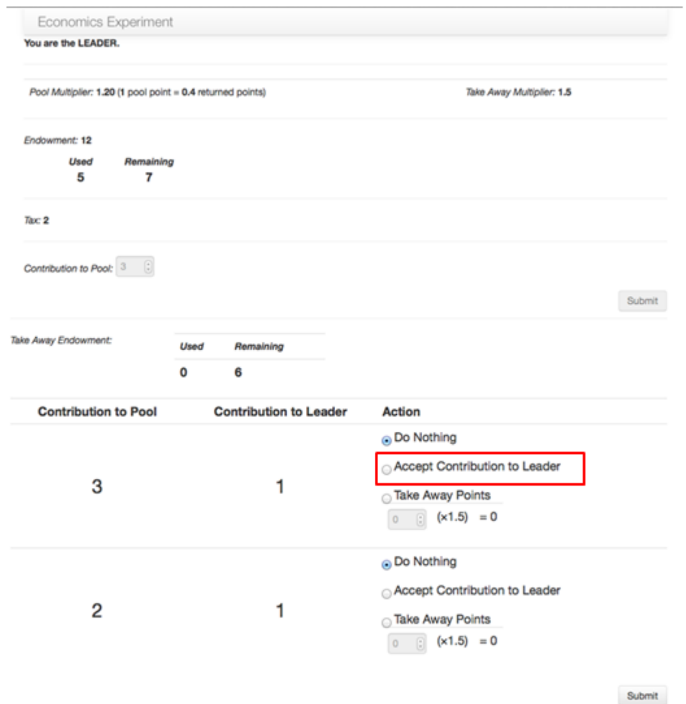
When you have made your decision about how many points to allocate to the group pool, click "Submit". After clicking "Submit", you will not be able to change your decision.



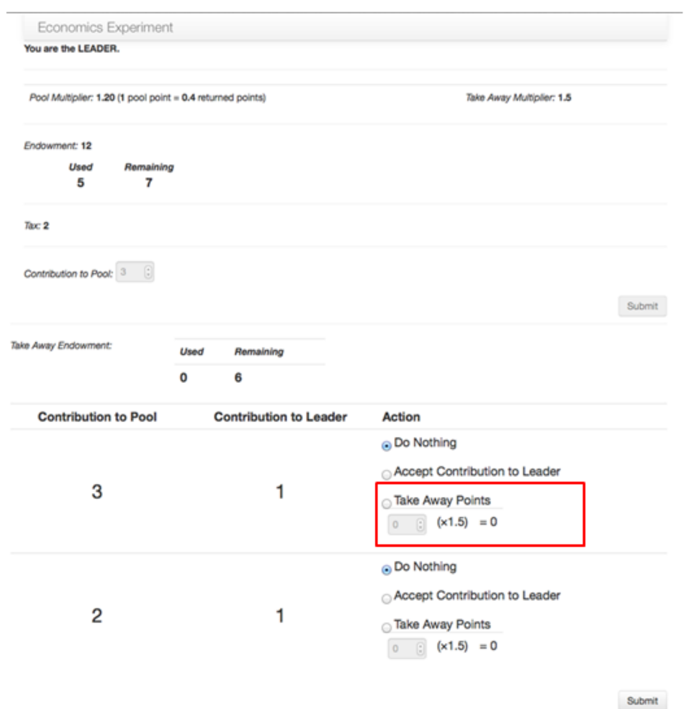
The actions of group members will appear below this after they make their decision. The Contribution column are the contributions to the group pool and the Payment column shows points given to you, the Leader.



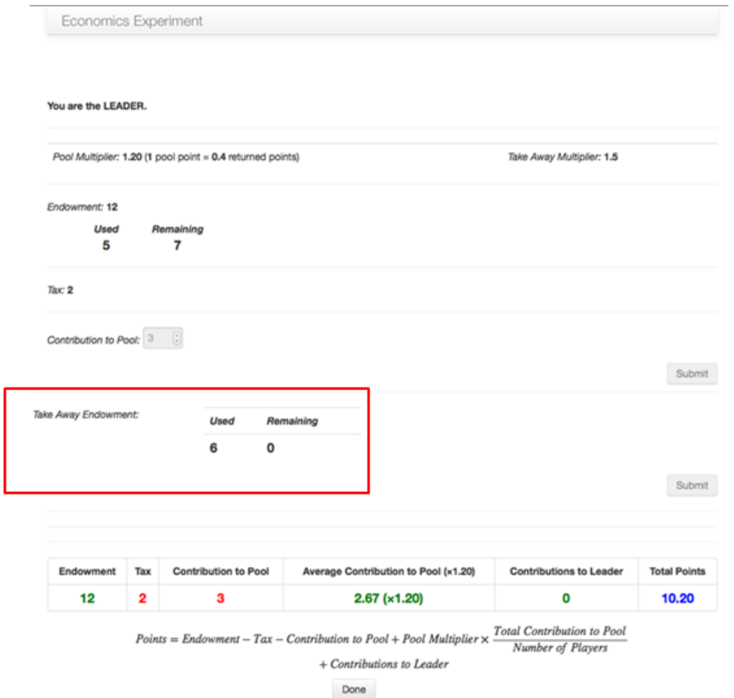
You may do nothing…



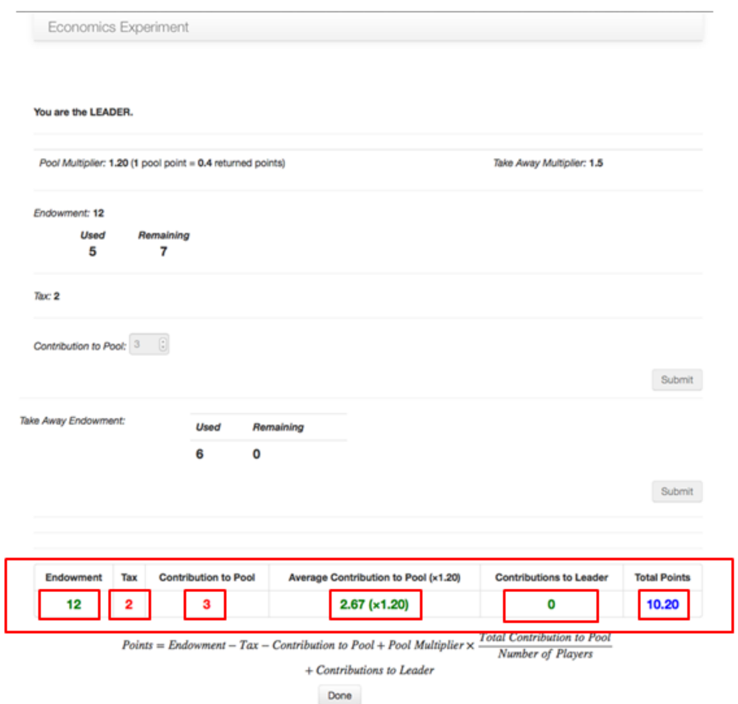
accept these points…



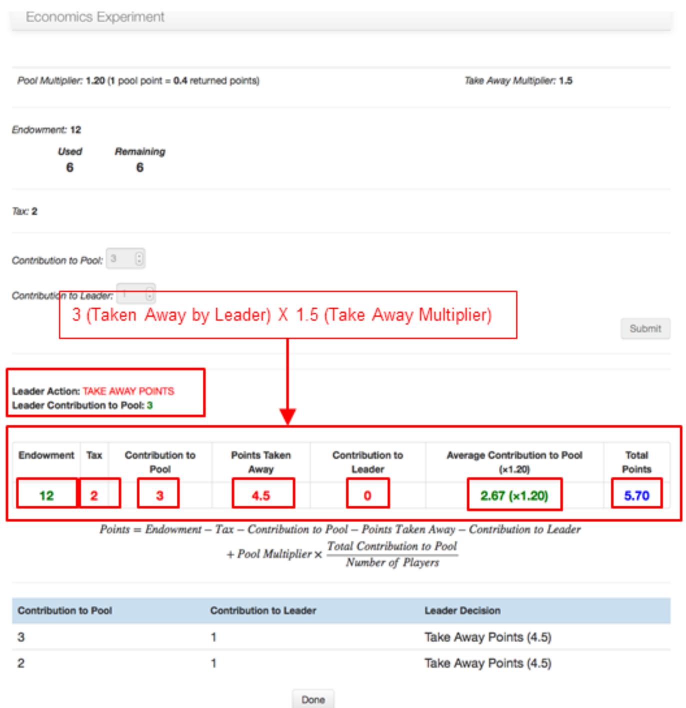
or take points away from any group member. If you take away points, you decide how much to take by allocating tax points to the group member. The total points taken away is calculated for you next to this box.



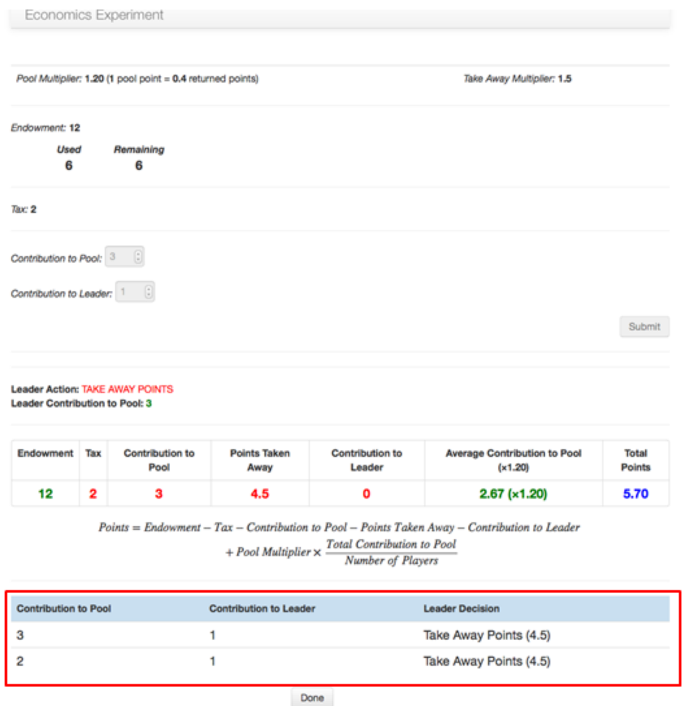
Any tax points not used for taking away points from group members will be returned to the experimenter. Similarly any points taken away from group members will also be returned to the experimenter.



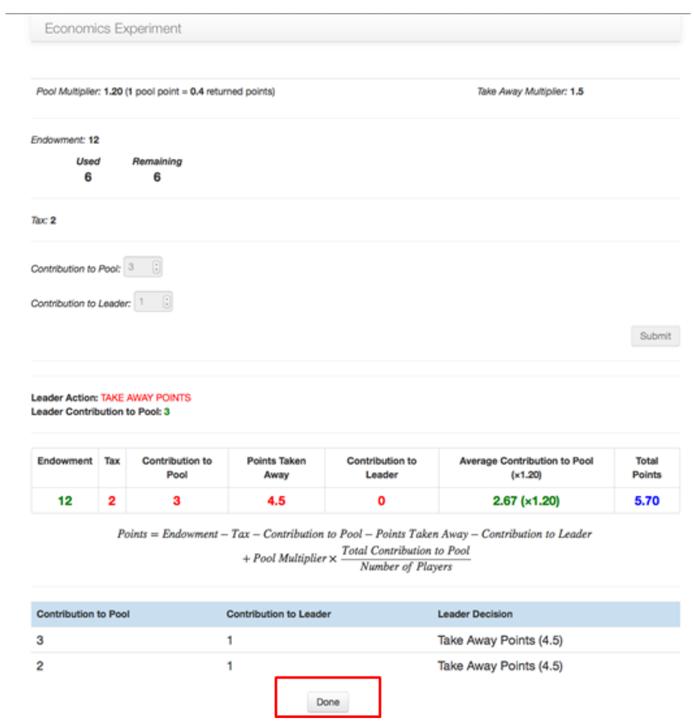
As a leader your earnings for this period will be calculated from your initial endowment minus your taxes, minus your contribution, plus the average contribution to the pool – 2.67, multiplied by the pool multiplier, plus any points you accepted, giving you a total of 10.2 for this period.



After the leader clicks submit, group members will be informed of the leader's action and the leader's contribution to the pool. The earnings of group members for this period will be calculated from their initial endowment minus their taxes, minus their contribution. Then either minus the points taken away by the leader OR minus the points given to the leader. Any points given to the leader and not accepted will be returned to the group member. In the example here, the leader chose to take points away from the group member. Finally, the group members share of the group pool will be added to this amount to calculate the total points for this period. Here the average contribution is shown – 2.67 points. This is multiplied by the pool multiplier and then added to give total points – 5.7 for this group member.



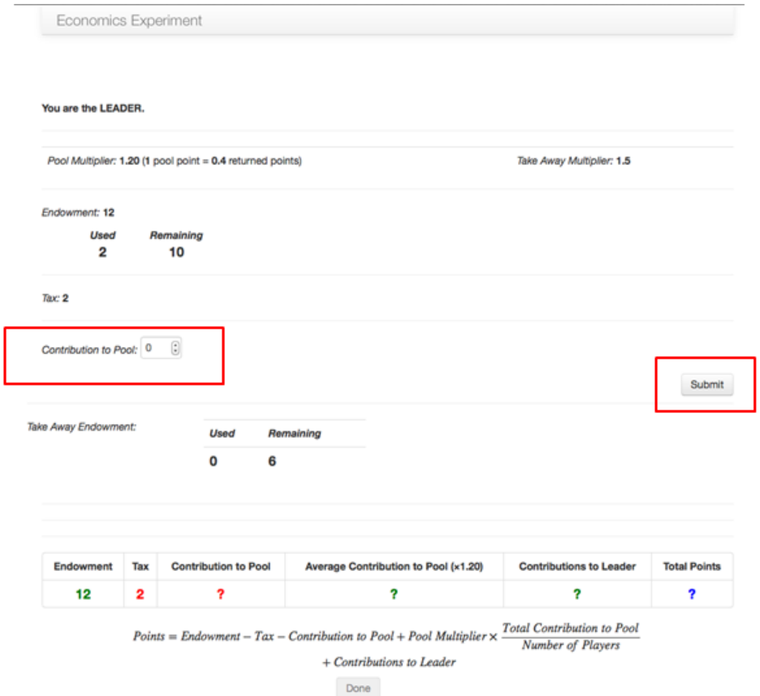
Group members will also be informed of the leader's decision for each decision made by other group members. Each line shows a group member's contribution to the pool, contribution to the leader and the leader's action. Group members themselves remain anonymous.



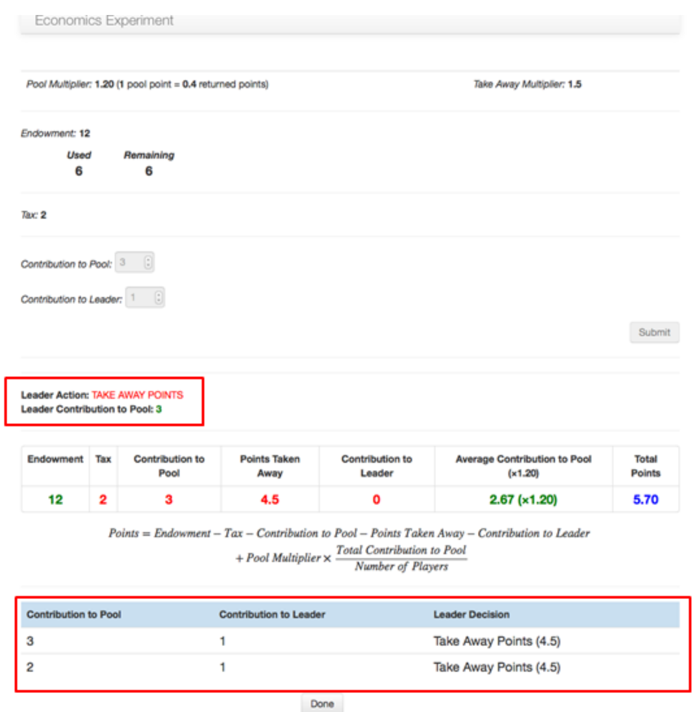
Please click "Done" at the bottom of the screen to move to the next period.

Please click "Continue" when you are ready to begin.

Subsequent Round



The difference between this round and previous rounds is that in this round, like other group members, the leader must choose how much to contribute to the group pool. They may also accept points from group members.



In addition, group members not only see the leader's action towards them, but also the leader's contribution to the group pool and the leader's actions towards other group members, including what the other group members contributed to the group pool and to the leader. As in all rounds, all participants are completely anonymous.

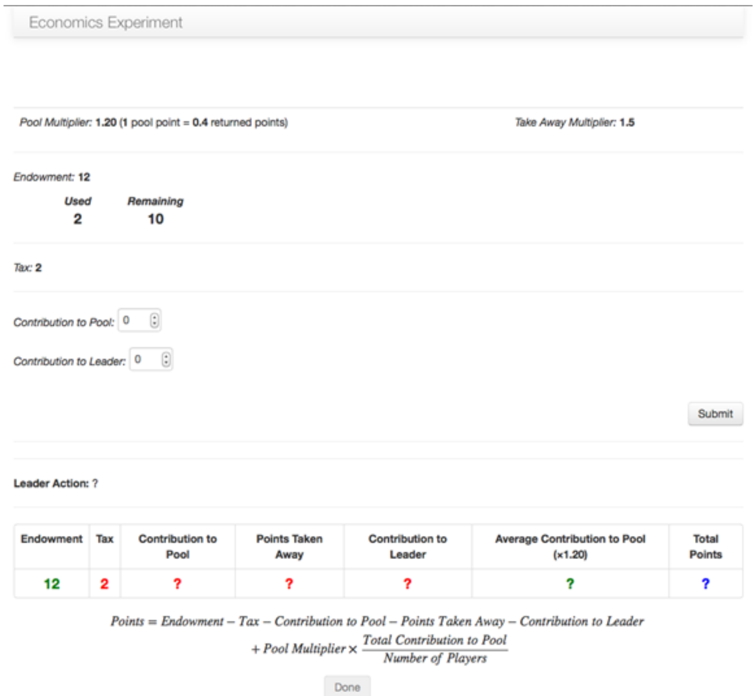
Please click "Continue" when you are ready to begin.

Quiz Questions

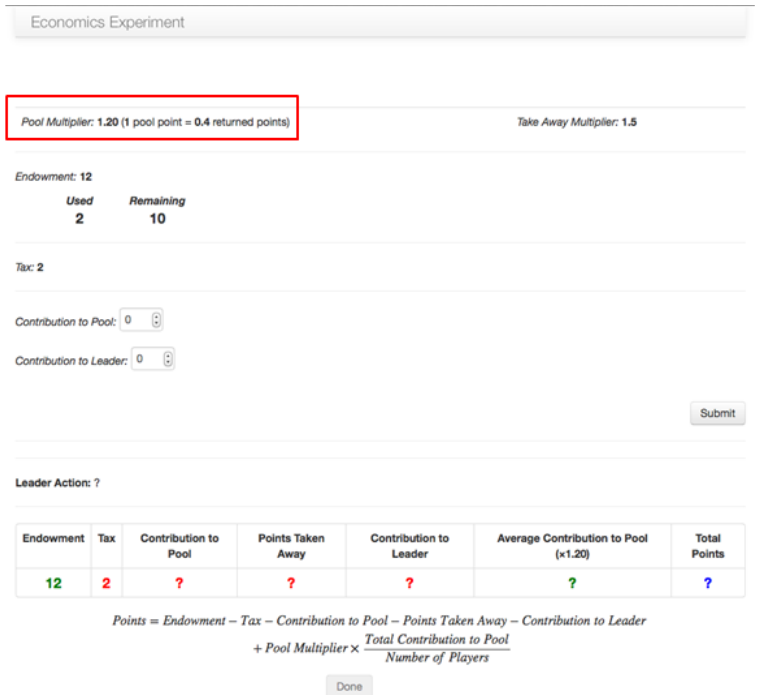
1. How many periods will you play?
    a. 10 periods
    b. More than 10 periods
    c. Less than 10 periods
    d. <u>An average of 10 periods, but possibly more and possibly less</u>
2. If everyone contributes the same amount to the group pool, will you:
    a. <u>Increase your points from your initial endowment</u>
    b. Decrease your points from your initial endowment
    c. Have the same points as your initial endowment
    d. Might increase your points or might decrease your points
3. How much does the leader contribute to the group pool?
    a. <u>However much they choose to allocate to the group pool</u>
    b. 10 points
    c. Depends on the points given to the leader
    d. Depends on the group member's contributions
4. Which of the following can the leader NOT do?
    a. Nothing
    b. Accept points from a group member
    c. Take points away from a group member
    d. <u>Accept points and take points away from the same group member</u>
5. If a leader does not accept points where do these points go?
    a. Goes to the leader anyway
    b. <u>Goes back to the group member</u>
    c. Goes to the experimenter
    d. Is added to the group pool
6. What actions of other group members can group members see?
    a. Leader's action towards them
    b. Leader's contribution
    c. Leader's contribution and all actions
    d. <u>All group member and leader actions</u>

Bribery Game with Leader Investment

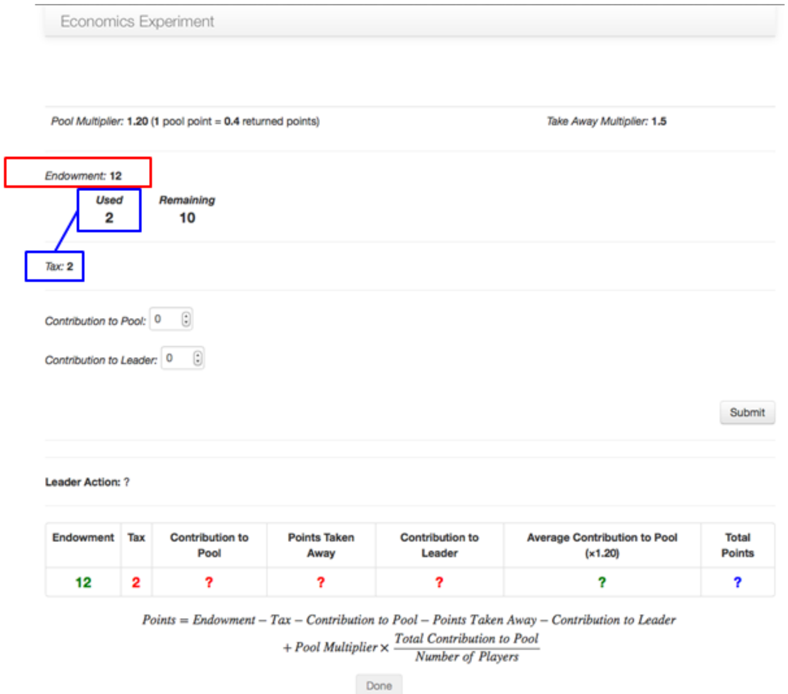
First Round



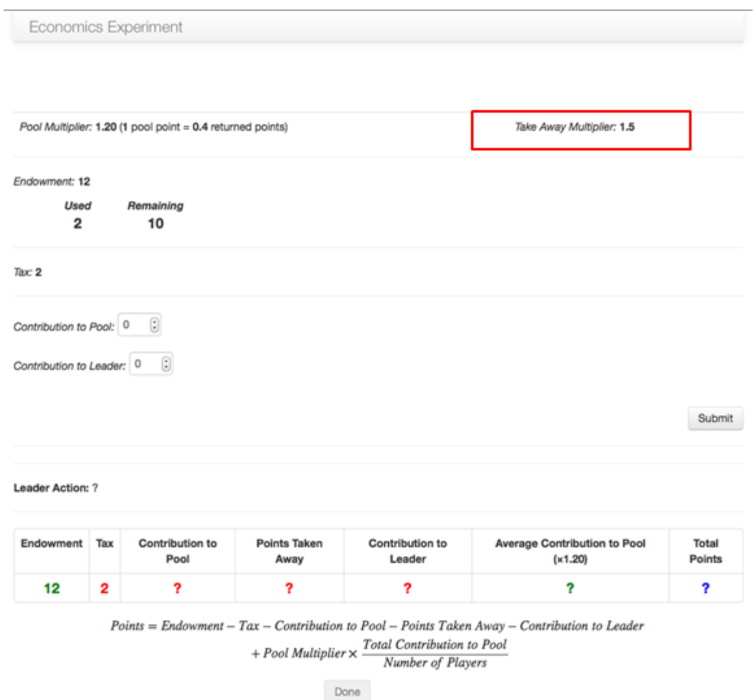
Welcome to the first round. The screen you see in front of you is the normal screen for play. In each period, one person from the group will be selected at random to play the role of a leader. The leader will have a slightly different screen. Please pay attention to all screens presented in these instructions as the instructions for other other rounds will only highlight the differences between that round and this one.



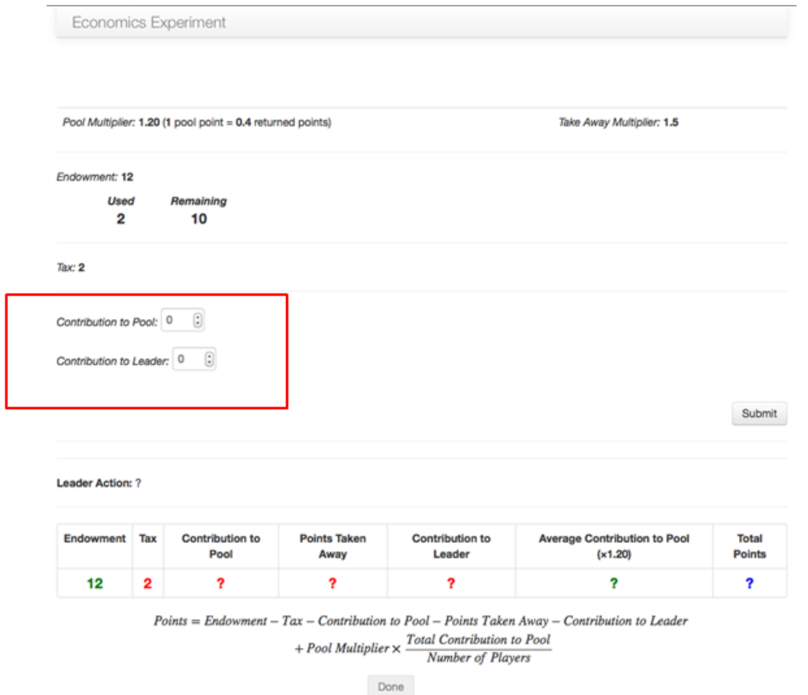
The first thing we'd like to draw your attention to is the pool multiplier. When group members allocate points to the group pool, the total allocated by the group is multipled by this number, the pool multiplier. In this case, it's 1.2. The new total is then distributed equally among everyone in the group, including the leader. This means that if every group member allocated all their points, all group members would increase their earnings by 20%. If no one allocates to the group pool, there's nothing to multiply by 1.2, and nothing to divide up. Another way to think about this is to consider how much each point in the pool is worth to each player. This is shown within parantheses to the right of the pool multiplier. In this case, with 3 players, each player gets 0.4 points for every point in the pool, no matter who contributes it. For example, if each player contributed 1 point pool, there would be 3 points in the pool and each player would get back 3 times 0.4 or 1.2 points.



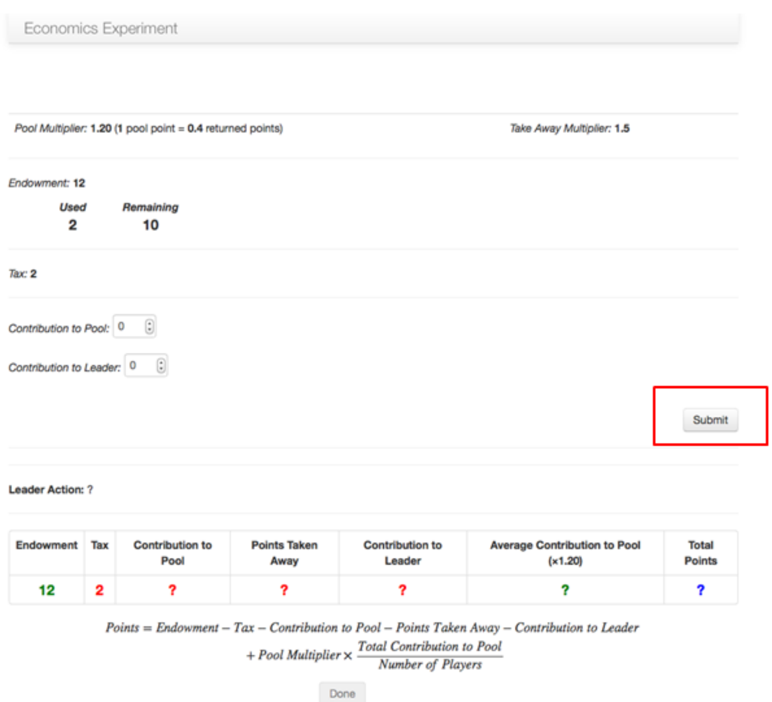
Below this, we see your initial endowment of 12 points. 2 points are taxed each period, which can be used by the leader to take points away from group members. This leaves you 10 points to allocate.



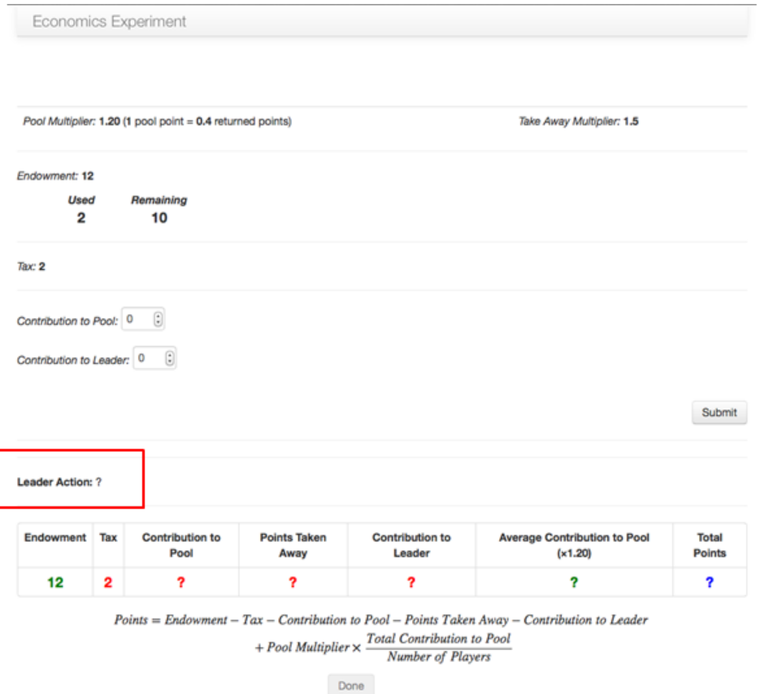
Here you see the Take Away multiplier. This is how much each tax point is multiplied by when the leader takes points away from a group member. For example, here, if the leader spent 1 tax point to take points away, 1.5 points would be taken from the targeted group member.



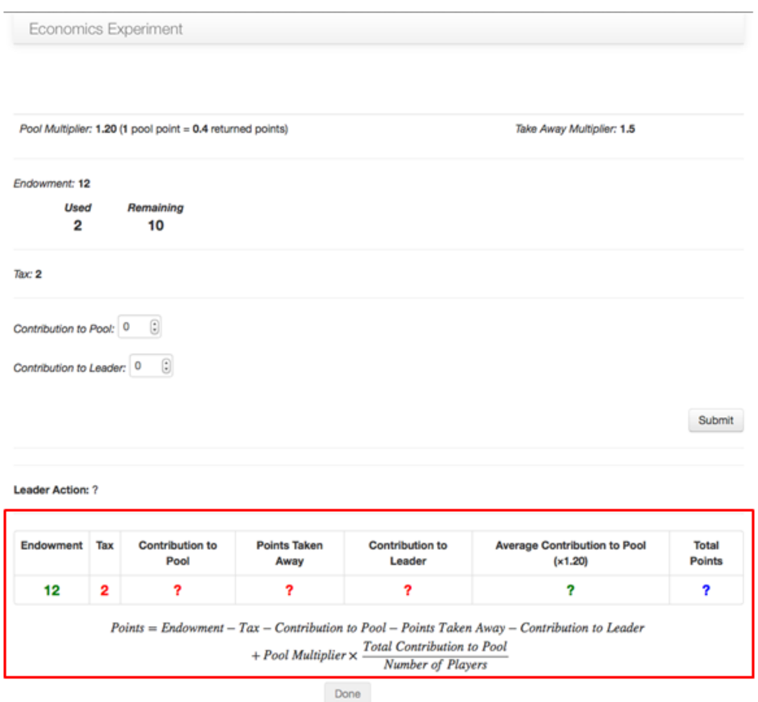
Here we see the boxes that are used to allocate your remaining 10 points to either the group pool or to the leader. Any points you do not allocate will be kept for yourself. When you hover your mouse over these boxes you will see an arrow to increase or decrease the amount in 1 point intervals or you may type in a value.



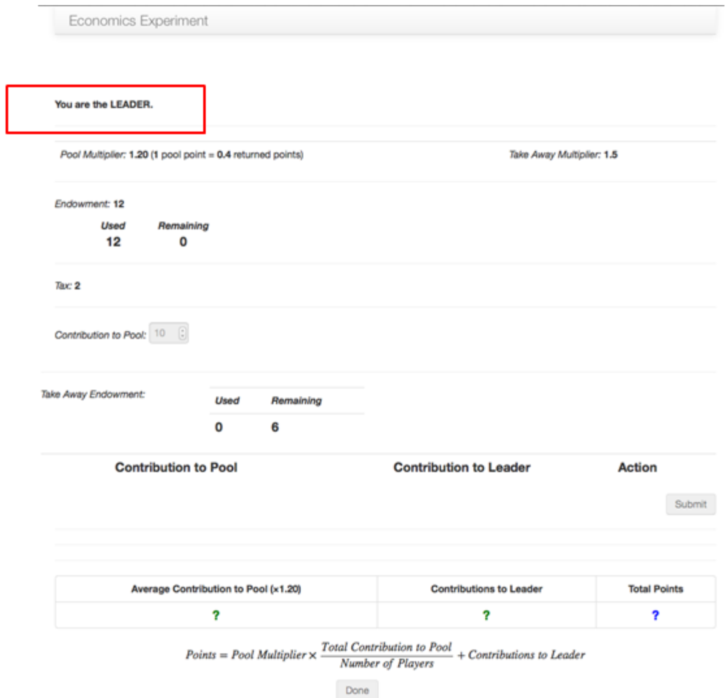
When you have made your decision, click “Submit”. After clicking “Submit”, you will not be able to change your decision.



This section will show you the leader's action. The leader has 3 choices – to take points from you, accept your points, or do nothing. They may only choose one of these options.



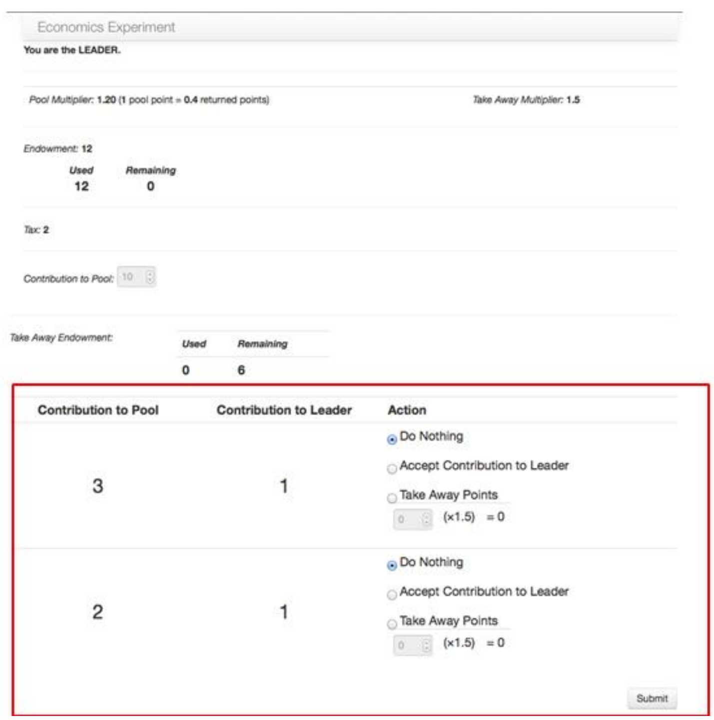
Here you see how your points for this period are calculated. Everything in green is added, everything in red is subtracted, and your total points for the period is shown in blue on the right. We will return to this after explaining the leader's screen.



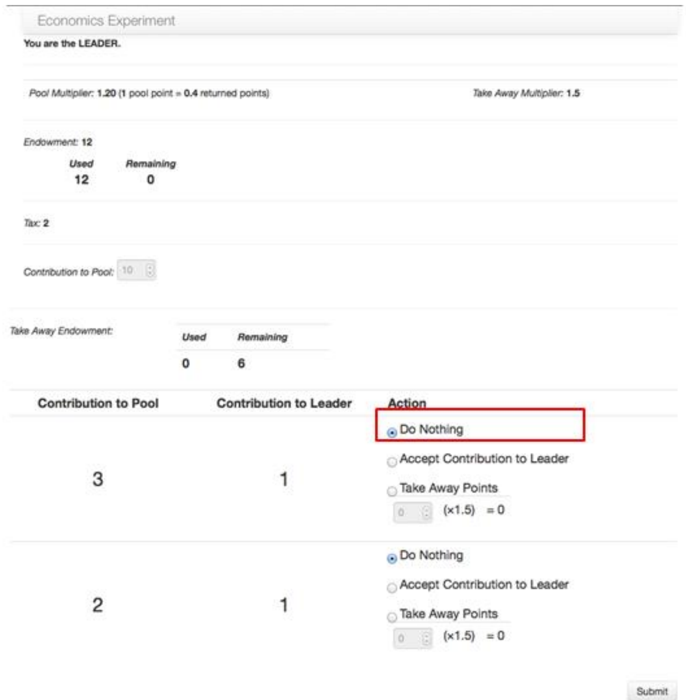
If you are the leader for a period, you will be identified at the top of the screen. Note that all group members, including the leader are anonymous to other group members. Your screen looks very similar to other group members, with a few differences.

Like other group members you pay a tax; but, unlike other group members you automatically contribute the maximum amount to the group pool.

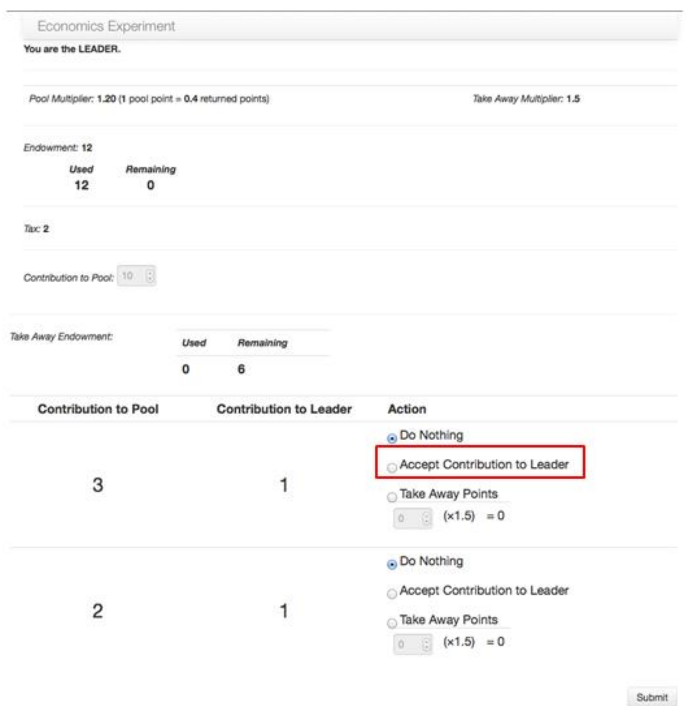
You also have access to tax points that you may use for taking away points from group members based on the Take Away multiplier.



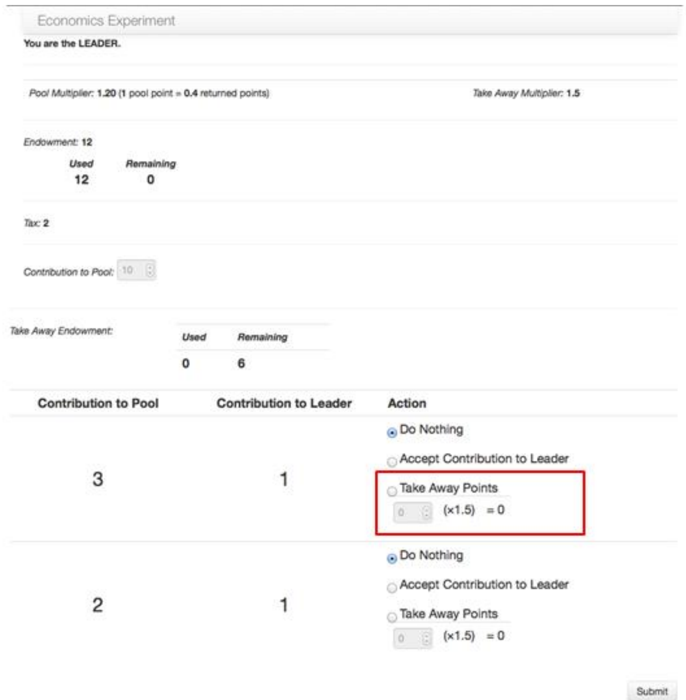
The actions of group members will appear below this after they make their decision. The Contribution column are the contributions to the group pool and the Payment column shows points given to you, the Leader.



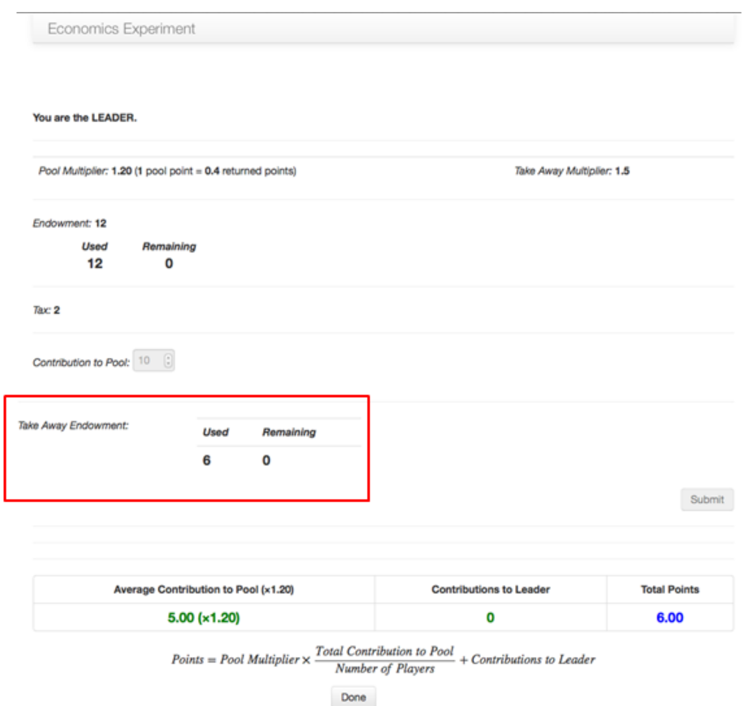
You may do nothing…



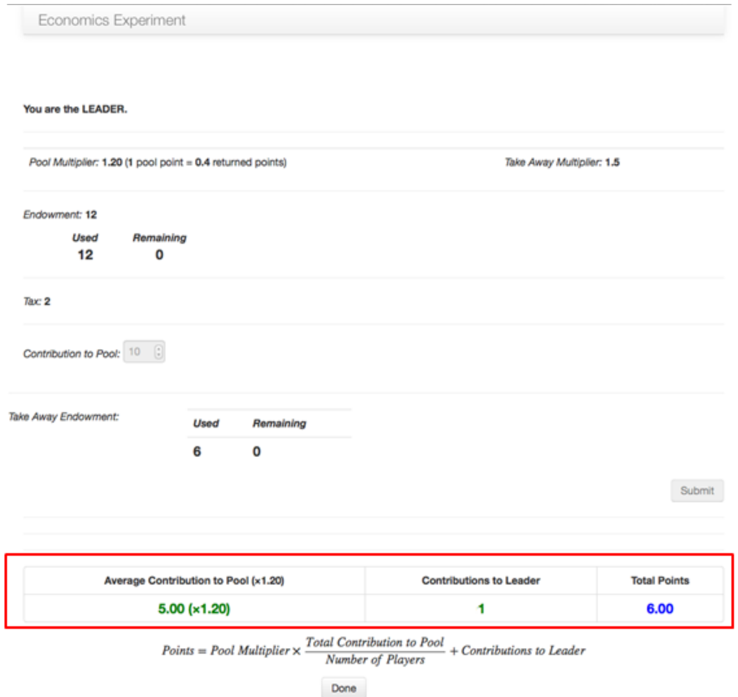
accept these points…



or take points away from any group member. If you take away points, you decide how much to take by allocating tax points to the group member. The total points taken away is calculated for you next to this box.

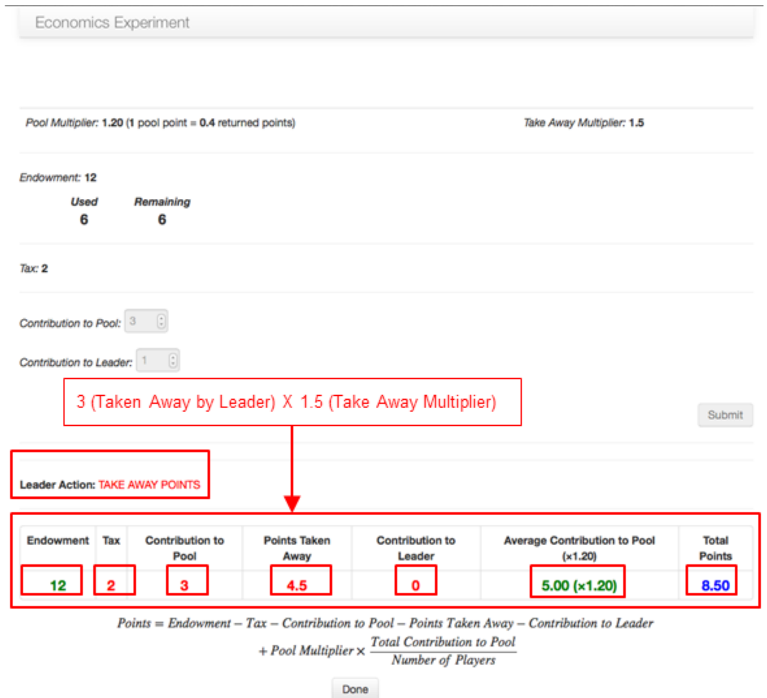


Any tax points not used for taking away points from group members will be returned to the experimenter. Similarly any points taken away from group members will also be returned to the experimenter.

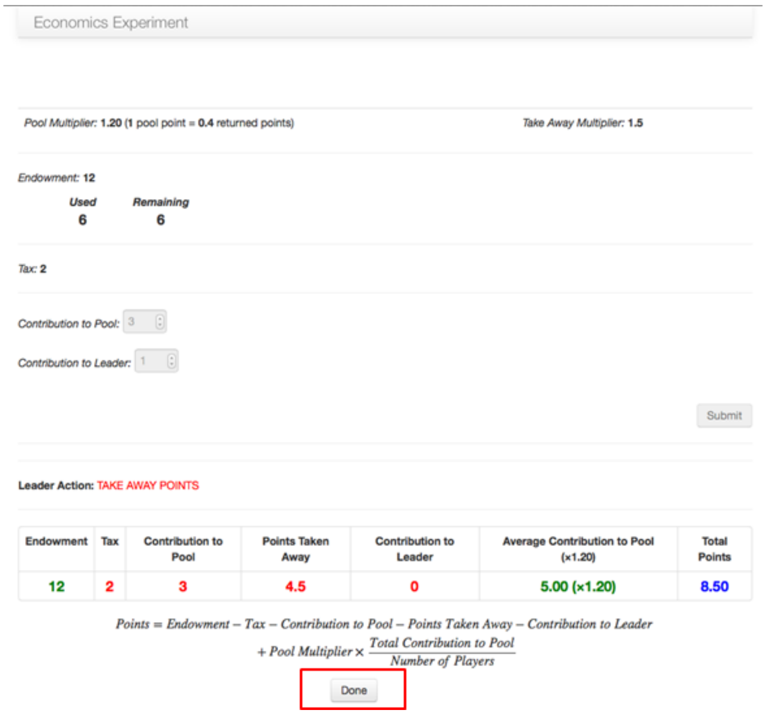


As a leader your earnings for this period are determined by how much you receive from your share of the group pool and the points you accept from each group member.

The average contribution of group members is shown here – 5 points. This is multiplied by the pool multiplier then added to accepted points to give the total points for this period – 6.



After the leader clicks submit, group members will be informed of the leader's action. The earnings of group members for this period will be calculated from their initial endowment minus their taxes, minus their contribution. Then either minus the points taken away by the leader OR minus the points given to the leader. Any points given to the leader and not accepted will be returned to the group member. In the example here, the leader chose to take points away from the group member. Finally, the group members share of the group pool will be added to this amount to calculate the total points for this period. Here the average contribution is shown – 5 points. This is multiplied by the pool multiplier and then added to give total points – 8.5 for this group member.

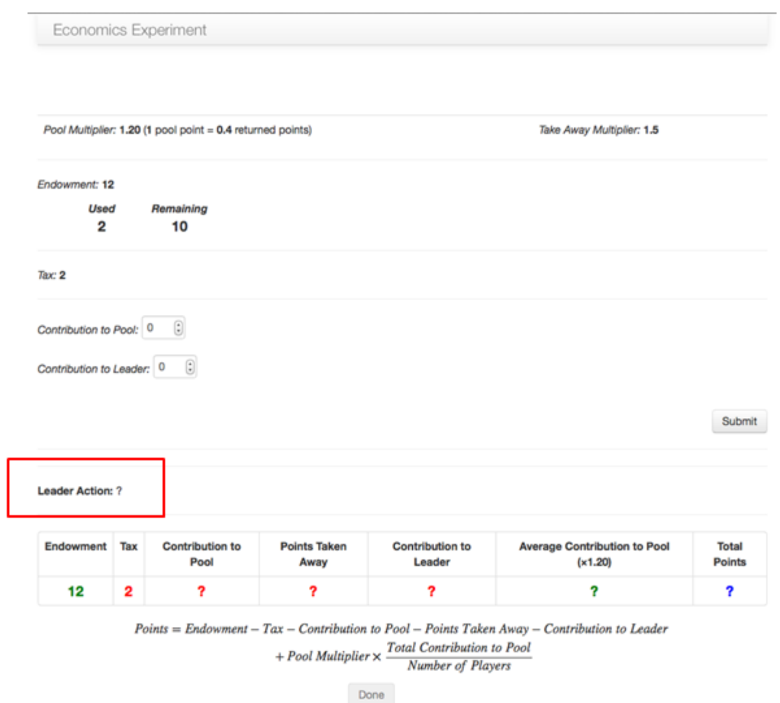


Please click "Done" at the bottom of the screen to move to the next period.

Please click "Continue" when you are ready to begin.

Subsequent Round

The difference between this round and previous rounds is that in this round, the leader automatically contributes the maximum amount to the group pool and may accept points from group members.



In addition, group members only see the leader's action towards them. They do not see the actions

of others group members or the leader's actions towards other group members. As in all rounds, all participants are completely anonymous.

Please click "Continue" when you are ready to begin.

Quiz Questions

1. How many periods will you play?
    a. 10 periods
    b. More than 10 periods
    c. Less than 10 periods
    d. <u>An average of 10 periods, but possibly more and possibly less</u>
2. If everyone contributes the same amount to the group pool, will you:
    a. <u>Increase your points from your initial endowment</u>
    b. Decrease your points from your initial endowment
    c. Have the same points as your initial endowment
    d. Might increase your points or might decrease your points
3. How much does the leader contribute to the group pool?
    a. However much they choose to allocate to the group pool
    b. <u>10 points</u>
    c. Depends on the points given to the leader
    d. Depends on the group member's contributions
4. Which of the following can the leader NOT do?
    a. Nothing
    b. Accept points from a group member
    c. Take points away from a group member
    d. <u>Accept points and take points away from the same group member</u>
5. If a leader does not accept points where do these points go?
    a. Goes to the leader anyway
    b. <u>Goes back to the group member</u>
    c. Goes to the experimenter
    d. Is added to the group pool
6. What actions of other group members can group members (not leader) see?
    a. <u>Leader's action towards them</u>
    b. Leader's contribution
    c. Leader's contribution and all actions
    d. All group member and leader actions

Theoretical Predictions

In this section we develop a cultural evolutionary framework in which to couch our understanding of bribe taking/giving and punishing behavior in a modified public goods game context. We can imagine two extreme approaches in understanding the behavior of participants in our experimental model. At one extreme, participant behavior is entirely governed by the parameters of the model. Their experience in the outside world, including internalized norms, personal preferences or learned behavior, have no bearing on their in-game behavior. At the other extreme, we can imagine participants whose behavior is unaffected by the experimental parameters. Their in-game behavior is entirely a product of their experience in the outside world. Between these two extremes, we can imagine participant behavior in our experimental model is a product of both participants' prior experience and the parameters of the experimental set up. Here, the experimental model captures how strong their prior norms and beliefs are relative to the strength of the experimental parameters. For example, in an experimental world of poor economic potential and weak leader punitive power, the effect of prior norms may be weaker—the experimental set up forces fewer high payoff behaviors. In contrast, when economic potential is poor, but leader power is strong, leaders may have an opportunity to select between different stable equilibria (stable depending on the strength of prior norms).

Our goal in developing this theoretical model is to understand these forces in general. That is, this theoretical model should be seen as complementing the experimental model; both the theory and experiment are attempting to model real-world structures and real-world behaviors. Inevitably, as with all models, we have to make some simplifying assumptions, but since both the theoretical and experimental model offer a window into the real world, they need not be the same simplifying assumptions - our theoretical model need not be constrained by experimental limitations since what we really want to understand is the real world, not the specifics of our in-game behavior. Thus, for example, in contrast to the experimental model, in our theoretical model, we assume a large population, such that leader contributions to the public pool do not affect the size of the pool in any meaningful way – prime ministers and presidents pay taxes, but these do not represent a large increase to Congressional coffers. Our overall goal is to understand how structural conditions, particularly economic potential and institutional or leader punitive power, affect corruption. We want to understand this both in terms of what equilibria different parameters allow, how these may shape norms, and the effect these norms may have on in-game behavior.

We take a cultural evolutionary approach[7-9] formally modeled using an adaptive dynamics evolutionary model[10,11*]. The gist of the adaptive dynamics approach is to hone in on cultural traits

[*] Cultural evolutionary theory (and the Dual Inheritance theoretical framework, more generally) emerged at the intersection of ecology, population biology, and anthropology[8,9] as an Ultimate-level[*12] explanation for the evolution of human psychology and human behavior. It describes the conditions that led our species to rely on socially learning from other members of the species and the conditions that led to this cultural knowledge, know-how, skills, beliefs, values,

(in this case bribery norms, which we can think of as behavioural dispositions or traits) and then analyze whether the following of these norms can constitute a set of stable self-reinforcing outcomes. We do this by looking to see whether small deviations from a posited equilibrium behavior (i.e. slightly more or slightly less bribery than the posited equilibrium bribery leads to higher or lower fitness), consistent with cultural learning, will lead back to the equilibrium trait[†]. If a trait is convergent stable, then fitness cannot be improved by a small deviation in either direction. The idea being that, when this is true, cultural evolutionary processes do not lead a population to diverge away from such behavior, hence suggesting it as an equilibrium.

In our adaptive dynamics model, we look at the evolution of the behavior of a population of citizens and a population of leaders over multiple generations. We are interested in where evolution will lead the behavior of these two populations, who are co-evolving within a particular set of economic potential and leader punitive power parameters. These populations will co-evolve since leader norms are contingent on citizen player norms and citizen player norms are contingent on leader norms. This set up can be interpreted as either citizens and leaders being different people or equivalently and more realistically, individuals behaving differently as ordinary citizens vs leaders (or citizens in charge of institutions). The argument we're making is similar to an Animal Farm allegory – the structural conditions, defined by economic potential and the punitive power of leaders will, over time, inevitably lead to different behaviors in these two roles.

Both the experimental and theoretical models are repeated one-shot interactions. In the experiment, this is because we randomly select a new leader each round to prevent any reputational effects or explicit conditioning on one leader's behavior. In principle, all rounds could have been played in parallel. Nonetheless, this repeated one-shot interaction does not imply that behavior is only a product of past experiences nor only a product of the in-game parameters. Instead, player behavior is a product of both the norms they have internalized from their past experiences (such as the prevalent norms in the countries they've lived in) and the parameters of the experiment. In the theoretical model, they are shaped by the co-evolution of the two populations adapting to the economic and leader parameters.

---

preferences, norms, and other aspects of behavior accumulating generation by generation to the point where no human could recreate this package. In one sentence, the answer is that humans tend to selectively learn from others (selective learning biases[13]), often copying behaviors and internalizing beliefs without fully understanding causality or payoffs (high fidelity transmission[14,15]). A cultural evolutionary approach has offered insights into several fields, including anthropology[16,17], evolutionary biology[18,19], economics[7,20-24], and psychology[25,26].

† Within a cultural evolutionary framework, off-equilibrium behavior can persist for some time, since individuals are reliant on social learning and since knowledge is imperfect and not all behavior is visible, individuals may not see equilibrium behavior. The off-equilibrium behavior may lead to a higher mean pay-off leading to higher societal outcomes. For example, although most tax authorities lack the funding to persecute a significant proportion of the population were they not to pay taxes, since most people see others around them paying taxes, they too will pay taxes despite the low probability of getting caught. One outcome of this process is the paradox of more cooperative societies displaying less punishment. This also provides a mechanism for moving between equilibria. For more information, see work on cultural-group selection[7,27].

The process for analyzing the cultural evolutionary adaptive dynamics involves setting up a fitness function, in this case equivalent to payoffs (norms and behaviors that lead to higher payoffs are more likely to culturally spread via selective copying, such as a success bias[28] and possibly, though not necessarily, genetically spread via number of offspring). We then calculate the equilibria for this fitness function by looking at when the first derivative is equal to zero. Next, we want to know whether an individual who deviates slightly from this equilibria can invade (moving the population to a different equilibrium). To do this, we first specify an invasion fitness function f—the relative fitness of a rare cultural mutant/deviant, calculated by subtracting the fitness of the resident equilibrium population from the invader fitness. We then calculate a selection gradient by taking the derivative of the invasion fitness f with respect to mutant trait (e.g. someone with a higher or lower contribution norm) and evaluating this derivative at the resident equilibrium values. If this value is negative, an invader with a lower value can invade; is positive, an invader with a higher value can invade. If zero, then we will calculate the second derivative to know if this is a stable equilibrium. If the second derivative is negative, then the value is a convergent stable evolutionary stable strategy (ESS). For those unfamiliar with this approach, it may be helpful to use a physical analog—distance, speed, and acceleration (or more accurately, displacement, velocity, and acceleration). The derivative of distance over time (metres) is speed (metres per second). The second derivative (derivative of speed) is acceleration (metres per second per second). The adaptive dynamics approach is the equivalent of looking at when an object is stationary (i.e. speed—derivative of distance—is 0) and confirming that these "equilibria" stationary points are convergent by confirming that objects decelerate around these points (i.e. acceleration—second derivative—is negative). If the second derivative were positive, objects would increase speed and move away from this stationary point, or in the present case, there would be positive selection for mutants away from this equilibrium.

Cultural Evolutionary Model

Here we formalize corrupt behavior using an adaptive dynamics evolutionary model[10,11] based on the Bribery Game. We assume a large population of citizens repeatedly playing the game with an indefinite time horizon over multiple generations. In our experiment, we had an experimental group that could fit in a room, and although we attempted to model an indefinite time horizon using Sampling

Based on a pilot of 50 participants, we expected a fairly large effect size and aimed to have at least 250 participants. All data was included, unless the software crashed (resulting in data loss or less than 10 rounds for that treatment). This happened for 5 of the 224 treatment groups. Incomplete data (and obviously lost data) was discarded prior to any analyses.

Block Randomization (see Experimental Design), we had a single generation and usually around 10 rounds—repeated one-shot interactions. In our cultural evolutionary model, we also assume a population of leaders or institutional norms (note that both citizen and leader populations are really a population of cultural traits, which are of course instantiated or possessed by people, but in reality we are tracking the evolution of these traits; i.e. in contrast to a genetic model, people don't have to die, they just have to change their beliefs and behaviors). That is, either these are different people or if they are the same people, their behaviour adapts depending on their role. The mapping from model to experiment is as follows: In the cultural evolutionary model, we identify the evolved equilibrium citizen and leader/institution normative behaviors that maximize fitness (in terms of payoff). In the experiment, we assume players (a) bring norms based on their ethnic background and direct 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. So, the model provides predictions about when we should see the economic potential and leader punitive power parameters affect contributions and bribe behavior and in what direction. We perform the following analyses:

1. We begin by analyzing the institutional punishment PGG with a fixed tax rate. We fixed the tax rate to more realistically capture real world institutions, where taxes and punishment are not directly correlated and where leaders can use the punitive powers of the state without a large personal cost (since there own taxes are a small part of the taxes contributing to the pool punishment or institution).
2. We then introduce the Bribery Game (BG) modification, whereby players have the option to offer bribes to the leader and players have the option to accept these bribes.

Let us begin by laying out our parameters and variables.

Parameters and Variables

Parameters are capitalized. Evolving variables are lower case.

Parameter	Description	Value in Experiment
E	Endowment	1.2
T	Taxes	0.2
M	Economic Potential (Multiplier on Public Good)	> 1. In experiment, the Marginal per capita rate of return (MPCR) is set at $[0.3, 0.6]$. The MPCR is M/N
S	Strength of leader (Multiplier on Leader Punishment)	> 1. In experiment, set at $[1, 3]$
N	Number of players	> 2

Variable	Description	Value
c_i	Contribution of player i to Public Good	$[0, E - T]$
b_i	Bribe of player i to Leader	$[0, E - T]$
p_i	Tax dollars assigned to punish player i	$[0, T \cdot N]$
α_i	Player i propensity to punish conditioned on size of contribution (when acting as Leader).	$[-\infty, \infty]$ >0 values indicate punishment for higher contributions <0 values indicate punishment for lower contributions
t_i	Player i threshold for 50% punishment conditioned on size of contribution (when acting as Leader)	$[0,1]$ Higher values indicate less punishment for higher values. Lower values indicate more punishment for higher values.
β_i	Player i propensity to punish conditioned on size of bribe (when acting as Leader)	$[-\infty, \infty]$ >0 values indicate punishment for higher bribes <0 values indicate punishment for lower bribes
h_i	Player i threshold for 50% punishment conditioned on size of bribe (when acting as Leader)	$[0,1]$ Higher values indicate less punishment for higher values. Lower values indicate more punishment for higher values.
F_i	Fitness of player i (equivalent to payoff in this model)	Function of player and leader behavior.

Note that we are treating economic potential and the strength of leaders as parameters, because we are interested in the effect of these structural factors on player behavior. However, for this to a be a more complete Ultimate-level cultural evolutionary model, we would need an explanation for what causes the value of these parameters, and how player norms affect these parameters. Although the endogenization (or at least explanation) for these parameter values is beyond the scope of the present work, we plan to explore this in future work with a focus on models of leadership.

Standard Institutional Punishment Public Goods Game (IPGG)

We begin with the standard institutional punishment PGG (IPGG) without any bribery. We can easily see that norms for:

(a) contributions (c) will tend toward zero without punishment
(b) levels of contributions are contingent on the strength of leaders (punishment multiplier; S) and tendency for leaders to punish contributions (dependent on α and t), and
(c) leaders will use taxes to punish, since these are not personally costly and since punishing increases the leader's payoff by increasing the size of the public good, which they share in.

We assume fitness and payoff are synonymous. Fitness (F_i) is given by endowment (E) minus taxes (T), contribution (c_i), and punishment ($S \cdot p_i$), plus the sum of all other contributions multiplied by the MPCR (${}^{M}/_{N}$):

$$F_i = E - T - c_i - S \cdot p_i + {}^{M}/_{N} \cdot \sum c_j$$

E and T are fixed, so:

$$F_i = 1 - c_i - S \cdot p_i + {}^{M}/_{N} \cdot \sum c_j$$

Next, we define the punishment assigned to player i as a function of the leader L's propensity to punish (α_L) and player i's contribution. To flexibly describe punishment behavior contigent on contributions, we use a logistic curve to describe this relationship, such that:

$$p_i = \frac{1}{1 + e^{-\alpha_L (c_i - t_L)}}$$

We illustrate this function in the figure below for different values of α and t, where t is the threshold contribution for punishment. Negative α indicates higher punishment for lower contributions (i.e. prosocial punishment), where more negative α indicates a steeper (more punitive) slope. Positive α indicates higher punishment for higher contributions (i.e. antisocial punishment), where more positive α indicates a steeper (more punitive) slope.

The threshold t determines the rate at which 50% of the punishment taxes are assigned. Lower t indicates a lower cutoff (e.g. if $\alpha < 0$, less tolerance for smaller contributions). Higher t indicates a higher cutoff. In the case when α is negative, this indicates more tolerance for smaller contributions.

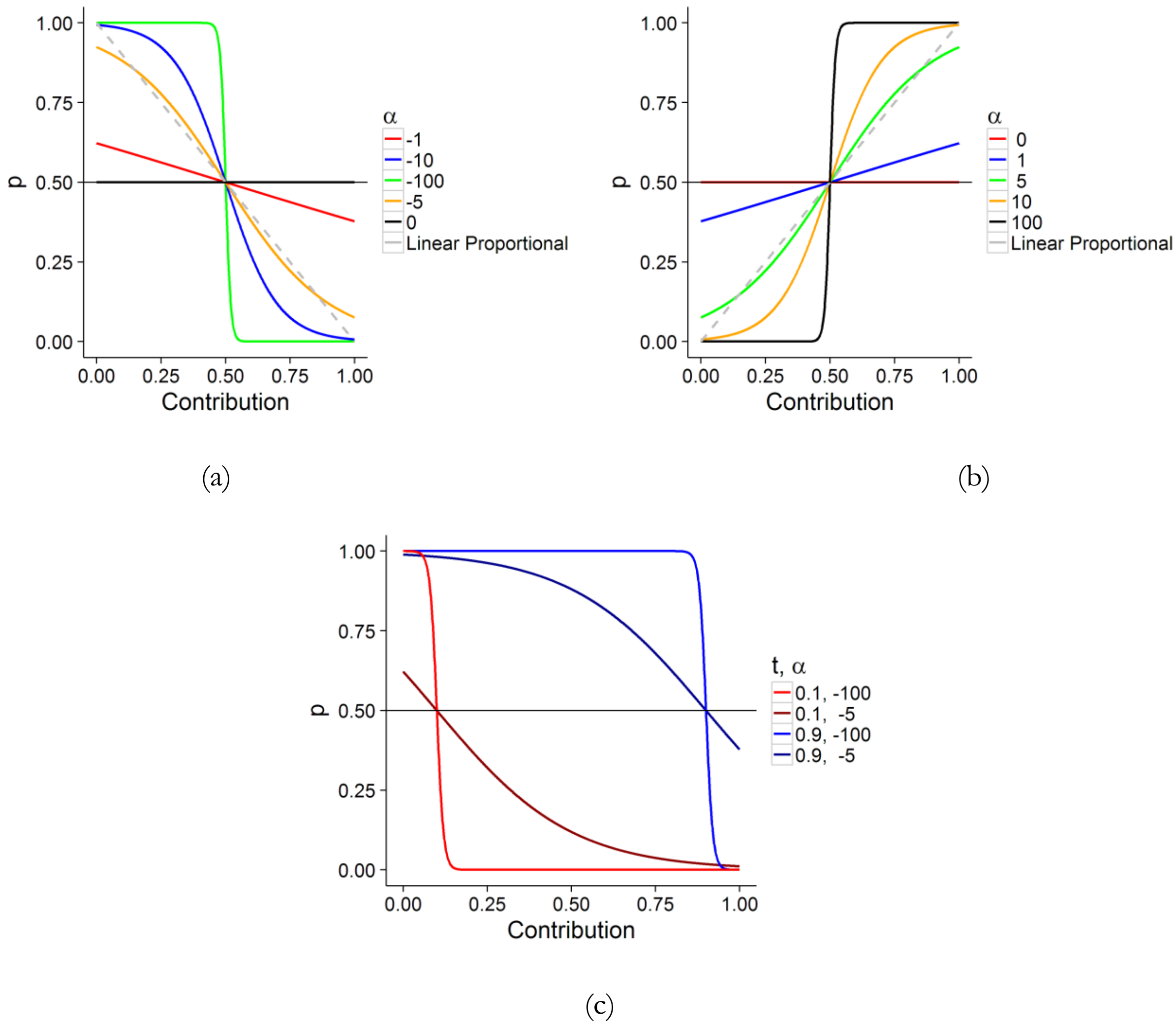


Figure S15. (a) Different negative values of α with a threshold of $t = 0.5$. i.e. larger punishments for smaller contributions. (b) Different positive values of α with a threshold of $t = 0.5$. i.e. larger punishments for larger contributions. (c) Negative values of α with extreme thresholds t. When $t = 0.1$ and α is large and negative (-100), there is a very large punishment for contributions less than 0.1 and almost no punishment for contributions more than 0.1 (almost a step function). In contrast, when $t = 0.9$ and α is large and negative (-100), there is a very large punishment for contributions less than 0.9 and almost no punishment for contributions more than 0.9 (again, almost a step function). Thus, by adjusting α and t, we can capture a great range of Leader punitive preferences.

Substituting p_i into F_i, payoff then becomes:

$$F_i = 1 - c_i - S \cdot \frac{1}{1 + e^{-\alpha_L(c_i - t_L)}} + {}^{M}/_{N} \cdot \sum c_j$$

Where the variables with subscript L capture the punishment preferences of the player designated as the Leader.

We solve this analytically by performing an invasion analysis of a monomorphic resident population (denoted with subscript r). In this homogenous population, everyone has the same contribution and everyone has the same preferences for punishment. Thus:

$$F_r = 1 - c_r - S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}} + {}^{M}/_{N} \cdot \sum c_r$$

Since everyone makes the same contribution, we can simplify our function:

$$F_r = 1 - c_r - S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}} + {}^{M}/_{N} \cdot N \cdot c_r$$
$$= 1 - c_r - S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}} + M \cdot c_r$$

Invader with a perturbed contribution

Let us now consider an invader (mutant) with a different contribution. That is, a player who deviates from the other players in how much they contribute to the public good. We denote this player with a subscript m. Without loss of generality, we will assume that the population is large enough so that the individual player's contribution doesn't significantly affect the size of the public good. That is:

$$Nc_r \approx (N-1)c_r + c_m$$

The growth rate $f_r(m)$ of the "mutant" (who offers a different contribution) player m in the resident r population of is given by:

$$f_r(m) = F_m - F_r$$
$$= 1 - c_m - S \cdot \frac{1}{1 + e^{-\alpha_r(c_m - t_r)}} + M \cdot c_r - \left(1 - c_r - S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}} + M \cdot c_r\right)$$
$$= -c_m - S \cdot \frac{1}{1 + e^{-\alpha_r(c_m - t_r)}} + c_r + S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}}$$
$$= c_r - c_m - S\left(\frac{1}{1 + e^{-\alpha_r(c_m - t_r)}} + \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}}\right)$$

Next, we find the selection gradient, by differentiating with respect to the mutant contribution and evaluating at the resident trait value $m = r$:

$$\frac{\delta f}{\delta c_m}|_{c_m = c_r} = -1 - \frac{\alpha_r S e^{-\alpha_r(c_m - t_r)}}{(e^{-\alpha_r(c_m - t_r)} + 1)^2}$$
$$= -1 - \frac{\alpha_r \, S \, e^{\alpha_r \, (c_m + t_r)}}{(e^{\alpha_r \, c_m} + e^{\alpha_r \, t_r})^2}$$

We see here that the resident contribution is irrelevant (it disappears). We can also see that if there were no punishment, the second part of this equation (the part after -1) would disappear and this derivative would always be negative. That is, a lower contribution would always be favored if leaders did not punish. Thus, equilibrium value is contingent upon punishment. We can look to see if higher contributions are ever favored (or if there is an equilibrium value) by looking at when this derivative is >0:

$$0 < -1 - \frac{\alpha_r\, S\, e^{\alpha_r\,(c_m + t_r)}}{(e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r})^2}$$

$$1 < -\frac{\alpha_r\, S\, e^{\alpha_r\,(c_m + t_r)}}{(e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r})^2}$$

$$(e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r})^2 < -\alpha_r\, S\, e^{\alpha_r\,(c_m + t_r)}$$

$$(e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r})^2 < -\alpha_r\, S\, (e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r})$$

$$e^{\alpha_r\, c_m} + e^{\alpha_r\, t_r} < -\alpha_r\, S$$

For this to be true, α_r must be large and negative (i.e. leaders must be more punitive toward lower values). Since the threshold, t_r and contributions c_m are both restricted to [0,1], we can simplify this function at look at it at the different values of t_r and evaluate c_m at the two extreme resident contributions of 0 & 1:

Assume: $c_m = 0; t_r = 0$

$$1 + 1 < -\alpha_r\, S$$

$$2 < -\alpha_r\, S$$

Assume: $c_m = 0; t_r = 1$

$$1 + e^{\alpha_r} < -\alpha_r\, S$$

Looking first at when contributions are zero, we see that a mutant with higher contribution norm can invade only when leaders have a stronger propensity to punish low contributors. For the same increase in contribution, we see that the punitive propensity can be less if their strength (i.e. punishment multiplier) is greater. Also, since more negative a_r values (i.e. $a_r \ll 0$) will cause e^{α_r} to tend toward zero, as you might imagine, a higher threshold for punishment (t) allows for less of a punitive slope (α; differential treatment of high vs low contributions) for this condition to be met.

Assume: $c_m = 1; t_r = 0$

$$e^{\alpha_r} + 1 < -\alpha_r\, S$$

Assume: $c_m = 1; t_r = 1$

$$2e^{\alpha_r} < -\alpha_r\, S$$

When resident contributions are maximal ($c_r = 1$), we see a similar pattern as before. The case when contributions are maximal and punishment thresholds are high is the case where a_r can be lowest and contributions sustained. Of course, since contributions are maximum, we should really look at when this condition is not met (i.e. look at when lower contributions can invade).

These analyses reveal that non-zero contributions can be sustained in the standard institutional PGG—even maximum contributions—as long as leaders punish lower values ($\alpha \ll 0$) and they are powerful enough to do so $S \gg 0$. As punitive preferences rise and leaders become more powerful, higher contributions can be sustained. Therefore, the stability of these contributions are contingent on a preference for punishment, which based on the norm literature[29], will depend on the norms the leader has experienced and is trying to uphold. Since leaders do not punish themselves and taxes are always extracted, we can assume that leaders are willing to punish, but we analyze the evolution of punitive preferences in the next section.

Perturbing Leader punitive preferences

As before, the payoff of a player i is given by:

$$F_i = 1 - c_i - S \cdot \frac{1}{1 + e^{-\alpha_L (c_i - t_L)}} + {}^{M}\!/_{N} \cdot \sum c_j$$

But since a Leader does not punish themselves, the fitness payoff for leaders (F_L) simplifies to:

$$F_L = 1 - c_L + {}^{M}\!/_{N} \cdot \sum c_j$$

It is trivial to show that a Leader is not incentivized to contribute (remember from before that the derivative is negative without punishment. Remember also that leader norms are different to player norms and subject to different structural conditions. Leaders experience no punishment), but their payoff is affected by the size of the public good, so their payoff is effectively:

$$F_L = 1 + {}^{M}\!/_{N} \cdot \sum c_j$$

We can re-arrange the player fitness in terms of this public good:

$${}^{M}\!/_{N} \cdot \sum c_j = F_i - 1 + c_i + S \cdot \frac{1}{1 + e^{-\alpha_L (c_i - t_L)}}$$

and substitute it in the Leader fitness:

$$F_L = 1 + F_i - 1 + c_i + S \cdot \frac{1}{1 + e^{-\alpha_L (c_i - t_L)}}$$

$$= F_i + c_i + S \cdot \frac{1}{1 + e^{-\alpha_L(c_i - t_L)}}$$

As before, we look for symmetric outcomes, with a mutant Leader (denoted with a subscript m) having different punishment preferences:

$$f_{Lr}(m) = F_{Lm} - F_{Lr}$$

$$= F_r + c_r + S \cdot \frac{1}{1 + e^{-\alpha_m(c_r - t_m)}} - \left(F_r + c_r + S \cdot \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}}\right)$$

$$= S \cdot \left(\frac{1}{1 + e^{-\alpha_m(c_r - t_m)}} - \frac{1}{1 + e^{-\alpha_r(c_r - t_r)}}\right)$$

We can then take the partial derivative with respect to α_m and t_m:

$$\frac{\delta f}{\delta \alpha_m}|_{\alpha_m = \alpha_r} = -\frac{S(c_r - t_m)e^{\alpha_m(c_r + t_m)}}{(e^{\alpha_m c_r} + e^{\alpha_m t_m})^2}$$

This is an implicit solution, however, since the punitive slope α_m is always on the exponent, regardless of other values, the derivative will always be negative, approaching 0 when $\alpha_m = -\infty$ (leaders become more punitive toward smaller contributions). A stronger leader (larger S) will make this a larger negative slope. The only other way for this derivative to be 0 (or positive) is if the resident contribution is equal the threshold t_m or below it (i.e. $c_r - t_m \leq 0$). So let's look at the partial derivative with respect to the threshold:

$$\frac{\delta f}{\delta t_m}|_{t_m = t_r} = \frac{\alpha_m S e^{\alpha_m(c_r + t_m)}}{(e^{\alpha_m c_r} + e^{\alpha_m t_m})^2}$$

Again, we have an implicit solution. However, here the sign of the derivative is entirely dependent on the sign of α_m. If $\alpha_m = 0$, the leader has no punitive preferences and all thresholds are irrelevant. If $\alpha_m < 0$, the threshold will tend toward the lowest value ($t_m = 0$) and thus Leaders will steeply punish non-contributors and be less punitive toward higher contributions. If $\alpha_m > 0$, the threshold will tend toward the highest value ($t_m = 1$) and thus leaders will punish maximum contributors, but will be less punitive toward lower contributions. So, t_m will either equal 0 or 1. But from the partial derivative with respect to α_m (i.e. $\frac{\delta f}{\delta \alpha_m}$), we know that α_m will always be negative, except when $c_r - t_m \leq 0$. If the threshold were 0, then $c_r \leq 0$, or really $c_r = 0$, since there can't be negative contributions. If the threshold were 1, then $c_r - 1 \leq 0$, which can only be true when $c_r = 1$. We are therefore left with the following situations:

1. Leaders are more punitive toward lower values, leading to higher contributions.
2. Contributions are maximum, contributions and threshold are equal, and Leader punitive values are irrelevant.
3. Contributions are zero, contributions and threshold are equal, and and Leader punitive values are irrelevant.

We want to know if any of these situations are convergent stable Evolutionarily Stable Strategies (ESS), such that minor deviations around these equilibria will inevibly lead back to these values. The cultural evolutionary process will over time lead societies to convergent stable ESS. To calculate if these are convergent stable ESS, we need to take the second derivative and look when it is less than 0 at these values:

$$\frac{\delta^2 f}{\delta \alpha_m^2} = \frac{S(c_r - t_m)^2 e^{a_L(c_r - t_L)}\left(e^{a_L(c_r - t_L)} - 1\right)}{(e^{a_L(c_r - t_L)} + 1)^3}$$

$$\frac{S(c_r - t_m)^2 e^{a_L(c_r - t_L)}\left(\boldsymbol{e^{a_L(c_r - t_L)} - 1}\right)}{(e^{a_L(c_r - t_L)} + 1)^3} < 0$$

This can only be negative when $e^{a_L(c_r - t_L)} < 1$.

$$\frac{\delta^2 f}{\delta t_m^2} = \frac{\alpha^2 S e^{a_m(c_r - t_m)}\left(e^{a_m(c_r - t_m)} - 1\right)}{(e^{a_m(c_r - t_m)} + 1)^3}$$

$$\frac{\alpha^2 S e^{a_m(c_r - t_m)}\left(e^{a_m(c_r - t_m)} - 1\right)}{(e^{a_m(c_r - t_m)} + 1)^3} < 0$$

This has the same requirement and can only be negative when $e^{a_m(c_r - t_m)} < 1$, which is true when $a_m(c_r - t_m) < 0$, which is met when $a_m < 0$ or $c_r - t_m < 0$.

Thus, cases 2 and 3 are not ESS strategies and only case (1) above applies. We can therefore conclude that leaders who are more punitive toward lower contributions will invade. Based on our invasion analysis of contributions, this means that contributions will increase. Moreover, from these analyses we can see that contributions will be higher when leaders are stronger (S is higher).

Do we find the same conclusion when leaders can accept bribes offered by players?

Bribery Game (BG)

The fitness functions in the BG are similar to the IPGG, but players have one additional choice and leaders can receive payoffs through a second channel. We show the following:

1. Players have no incentive to offer bribes, except if they will be punished for not doing so.
2. Leaders have a greater incentive to punish for lack of bribes than for lack of contributions
3. As in the IPGG, Leaders have a greater ability to impose their will when S is higher.

4. If players have a non-zero tendency to contribute (beyond punishment, for reasons not explicitly captured by this model, such as internalized norms), a Leader's incentive to punish for bribes will be slightly dampened when economic potential is higher (multiplier on public good, $\boldsymbol{M}$, is higher).

A player's fitness in the BG is given by:

$$F_i = 1 - c_i - \boldsymbol{b_i} - S \cdot p_i + {}^{M}/_{N} \cdot \sum c_j$$

And the Leader's fitness is:

$$F_L = 1 + {}^{M}/_{N} \cdot \sum c_j + \sum \boldsymbol{b_j}$$

Note the bolded $\boldsymbol{b}$ for the bribe. Note also that as in the IPGG, Leaders have no incentive to contribute, since they do not punish themselves.

The punishment can now be conditioned not only on the contribution, but also the bribe:

$$p_i = \frac{1}{1 + e^{-\alpha_L(c_i - t_L)}} + \frac{\boldsymbol{1}}{\boldsymbol{1 + e^{-\beta_L(b_i - h_L)}}}$$

There are two additional constraints that we are ignoring for now: (1) the percent punishment cannot exceed 100% (i.e. 1) and (2) $b_i + c_i \leq 1$. The player payoff or fitness functions then becomes:

$$F_i = 1 - c_i - b_i - S \cdot \left(\frac{1}{1 + e^{-\alpha_L(c_i - t_L)}} + \frac{1}{1 + e^{-\beta_L(b_i - h_L)}}\right) + {}^{M}/_{N} \cdot \sum c_j$$

From the function above, we can see that from the player's perspective, bribes and contributions are symmetric in terms of loss to endowment and potential loss via punishment. If anything players have even less of an incentive to offer a bribe than contribute, since there is no return on bribes, but there is at least the potential return from the public good for contributions. Thus, player behavior for bribes, as with contributions, are dependent on leader punishment behavior. Thus, we need to analyze the invasion of leaders with different punitive preferences:

Perturbing Leader punitive preferences

Leaders should be optimizing their fitness:

$$F_L = 1 + {}^{M}/_{N} \cdot \sum c_j + \sum b_j$$

Again restricting to a symmetric outcome, this becomes:

$$F_{Lr} = 1 + Mc_r + Nb_r$$

And players fitness become:

$$F_r = 1 - c_r - b_r - S \cdot \left(\frac{1}{1 + e^{-\alpha_r (c_r - t_r)}} + \frac{1}{1 + e^{-\beta_r (b_r - h_r)}} \right) + Mc_r$$

Since a Leader's punitive ability are symmetric with respect to contributions are bribes, the key is which punishment enhances their fitness (recall that there is a tradeoff between these punishment allocations). We can see that punishing for not giving sufficient bribes is always more fitness enhancing since $Nb > Mc$.

More precisely, it is always more fitness enhancing because the game is set up to embed a public goods dilemma (i.e. $M/_N < 1$). Thus, as in our analysis of the IPGG, Leaders are incentivized to punish, but this time, to punish low bribes, instead of low contributions. And again, this ability is greater when S is greater. Therefore, the logic here generates distinct predictions for strong leaders/institutions in the BG versus the IPGG: stronger leaders encourage more bribes when bribery is an option (BG), but more contributions to the public good when bribery is not an option (IPGG).

A Leader's payoff through bribes increases with the size of the population. For a Leader to be incentivized to punish contributions, at least one of the following must be true:

(a) The world needs to no longer be in a public goods dilemma (i.e. $M/_N \geq 1$) and it's individually advantageous to contribute to the public good—this may well be true in some real world cases, but is not captured in our game.

(b) Players must be more reluctant to offer bribes rather than contribute or have a non-zero tendency to contribute. This is possible since players do have a potential personal return on contributions (via the public good provisioning), but not on bribes, but could also be true if there is an exogenous norm for prosocial contributions, or an anti-corruption norm against offering bribes (when we experimentally model these dynamics, we measure proxies for corruption/anticorruption norms through exposure to these norms). Either of these factors tilt towards contributions and away from bribes. And the interaction with economic potential is as follows: such normative exogenous preferences are more likely to overcome the leader's payoff associated with bribes when economic potential is higher.

Predictions Summary

The logic laid out thus far leads to the following predictions:

1. For the regression on contribution: $c = \beta_1 S + \beta_2 M + \epsilon$:
    a. $\beta_1 > 0$ in IPGG, i.e. stronger leaders result in higher contributions
    b. $\beta_1 < 0$ if BG, i.e. stronger leaders results in lower contributions
    c. $\beta_2 > 0$ in IPGG

d. $\beta_2 \geq 0$ in BG (depending on prior contribution preferences not captured by our formal theory)

2. For the regression on bribes: $b = \beta_1 S + \beta_2 M + \epsilon$:
   a. $\beta_1 > 0$, i.e. more bribes offered when leaders more powerful
   b. $\beta_2 \leq 0$, i.e. no change in bribes or less bribes offered when economic potential is higher
3. In the BG, for the regression on punishment: $p = \beta_1 c + \beta_2 b + \beta_3 b \times S + \beta_4 b \times M + \epsilon$:
   a. $\beta_3 < 0$, i.e. more punishment will be allocated for bribes when leaders are more powerful.
   b. $\beta_4 \leq 0$, i.e. no change in punishment based on economic potential, but if there is a change, it will be less when economic potential is higher.
4. By corollary, for the regression on Leader decisions: $Accept\ Bribe = \beta_1 S + \beta_2 M + \epsilon$:
   a. $\beta_1 > 0$, i.e. more acceptance of bribes (compared to doing nothing or punishing) when leaders are more powerful.

These predictions should be treated with caution due to the underlying assumptions in the theoretical model and experimental model. For example, although both the experimental model and theoretical model are one-shot interactions, the behavior in the experiment is the product of both the parameters the participants experienced in the game and the norms they have brought into the game due to the parameters they have experienced in the real world. This theoretical model gives some insight into how these parameters shape norms in the real world and shape in-game behavior and the degree and the conditions in which each are likely to exert a stronger influence. Indeed, since our model suggests that bribes ought to the only channel for leaders when they are powerful enough to extract bribes, and yet we see contributions and punishment for contributions, these norms are indeed likely having some effect. We have attempted to disentangle the effect of these norms through our ethnic and experience corruption scores, but a proper test would involve running this experiment in different countries around the world (which we plan to do as a follow up).

Supplementary Results

Data Analyses

We analyzed our data using generalized linear mixed models (GLMM), calculating coefficients using a Monte Carlo-Markov Chain (MCMC) implemented by the R package *MCMCglmm*[30]. All Bayesian models pass the Gelman and Rubin[31] convergence diagnostic, implemented in the *gelman.diag* function of the *coda*[32] package. Categorical models are rescaled to log odds as per Hadfield[33] course notes for *MCMCglmm*. Confidence intervals (95%) were calculated as Highest Posterior Density (HPD) using the *HPDinterval* function in the *coda* package[32].

We provide a frequentist equivalent to each analysis (with no substantive difference in interpretation).

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.

All data is available at FigShare: https://doi.org/10.6084/m9.figshare.5004956

Predictions

Based on the logic that leaders can make more money by using money to extract bribes from every player than by increasing the size of the public good, we predict the following:

5. For the regression on contribution: $c = \beta_1 S + \beta_2 M + \epsilon$:
    a. $\beta_1 > 0$ in IPGG, i.e. stronger leaders result in higher contributions
    b. $\beta_1 < 0$ if BG, i.e. stronger leaders results in lower contributions
    c. $\beta_2 > 0$ in IPGG
    d. $\beta_2 \geq 0$ in BG (depending on prior contribution preferences)
6. For the regression on bribes: $b = \beta_1 S + \beta_2 M + \epsilon$:
    a. $\beta_1 > 0$, i.e. more bribes offered when leaders more powerful
    b. $\beta_2 \leq 0$, i.e. no change in bribes or less bribes offered when economic potential is higher
7. In the BG, for the regression on punishment: $p = \beta_1 c + \beta_2 b + \beta_3 b \times S + \beta_4 b \times M + \epsilon$:
    a. $\beta_3 < 0$, i.e. more punishment will be allocated for bribes when leaders are more powerful.
    b. $\beta_4 \leq 0$, i.e. no change in punishment based on economic potential, but if there is a change, it will be less when economic potential is higher.
8. By corollary, for the regression on Leader decisions: $Accept\ Bribe = \beta_1 S + \beta_2 M + \epsilon$:
    a. $\beta_1 > 0$, i.e. more acceptance of bribes (compared to doing nothing or punishing) when leaders are more powerful.

Variables

cContribution	Raw, centered contribution to public good
zContribution	z-score of contribution to public good
cBribe	Raw, centered bribe to Leader
zBribe	z-score of bribe to Leader
cPunishment	Raw, centered punishment
zPunishment	z-score of punishment
cPlayerExposureCorruption	Raw, centered player corruption score from countries they've lived in
zPlayerExposureCorruption	z-score of player corruption score from countries they've lived in
cLeaderExposureCorruption	Raw, centered leader corruption score from countries they've lived in
zLeaderExposureCorruption	z-score of leader corruption score from countries they've lived in
MPCR	Marginal per capita rate of return (0.3 or 0.6)
LeaderPower	Multiplier on leader punishment (1 or 3)
Cond	Treatment: Control=Public Goods Game; BG=Bribery Game; BG_Part_Trans = BG + Partial Transparency BG_Full_Trans = BG + Full Transparency BG_Leader= BG + Leader Investment
Period	Period of game – only first 10 were analyzed
Version	Depending on whether background questions were given before or after the game
Order	Each subject played 4 of 5 treatments. Order specifies the order in which they played that particular treatment
Subjects	Number of players in the group
Age	Age in years
Male	Gender (Male = 1; Female = 0)
PID	Participant ID
GroupNum	Group ID

Cost of Corruption

Here we compare behavior in the standard institutional punishment Public Goods Game to behavior in the Bribery Game – identical in all ways, except the additional option of the bribe. Here and in all cases, we show the R code for the model, with the output in a clean table format. The data and R code are available on DataDryad.

Contributions

Bayesian

Model

```
mcmcmodel <- MCMCglmm(cContribution ~ factor(MPCR)+factor(LeaderPower)
+ factor(Cond) + Period + factor(Version) + Subjects +
as.numeric(Order) + scale(age) + male, random=~PID:GroupNum,
data=levi[levi$Cond=="BG" | levi$Cond=="Control",])
```

For the standardized version, zContribution was regressed instead.

Results

	Coefficient	**95% CI**	**p-value**
Economic Potential	1.84	1.26, 2.50	$< .001$
Strong Leader	−0.41	−1.01, 0.20	.198
Bribery Game	−1.37	−1.54, −1.21	$< .001$
Period	−0.02	−0.04, 0.01	.142
Version	1.02	0.42, 1.61	$< .001$
Subjects	0.09	−0.21, 0.38	.560
Order	0.57	0.47, 0.68	$< .001$
Age	0.01	−0.30, 0.26	.976
Male	0.97	0.38, 1.61	.002
(Intercept)	−2.54	−4.12, −0.80	.004
Obs.		2716	
N		248	
Group Num.		56	
DIC		10650.22	

Table S1. MCMC GLMM regression on raw, unstandardized contribution.

	Coefficient	95% CI	p-value
Economic Potential	0.58	0.39, 0.78	< .001
Strong Leader	−0.13	−0.32, 0.09	.238
Bribery Game	−0.43	−0.49, −0.38	< .001
Period	−0.01	−0.01, 0.00	.188
Version	0.32	0.12, 0.52	.002
Subjects	0.29	−0.06, 0.11	.562
Order	0.18	0.15, 0.22	< .001
Age	0.00	−0.09, 0.08	.948
Male	0.31	0.12, 0.52	< .001
(Intercept)	−0.81	−1.35, −0.34	.002
Obs.		2716	
N		248	
Group Num.		56	
DIC		4394.94	

Table S2. MCMC GLMM regression on z-score of contribution to calculate standardized coefficients.

Frequentist

Model

```
model <- lmer(cContribution ~
factor(MPCR)+factor(LeaderPower)+factor(Cond) + Period +
factor(Version) + Subjects + Order +age+male+ (1| PID)+ (1 |
GroupNum), data= data=levi[levi$Cond=="BG" | levi$Cond=="Control",])
```

Results

	Coefficient	95% CI	p-value
Economic Potential	1.95	0.86, 3.04	.001
Strong Leader	−0.31	−1.41, 0.80	.602
Bribery Game	−1.39	−1.56, −1.23	< .001
Period	−0.02	−0.04, 0.01	.144
Version	1.32	0.29, 2.32	.013
Subjects	0.08	−0.47, 0.64	.777
Order	0.54	0.43, 0.65	< .001
Age	−0.01	−0.05, 0.04	.851
Male	0.72	0.28, 1.17	.002
(Intercept)	−2.69	−5.83, 0.46	.109
Obs.		2716	
N		248	
Group Num.		56	
R^2		0.76	

Table S3. Multilevel model regressing raw, unstandardized contribution with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.76$.

Standardized Score

	Coefficient	95% CI	p-value
Economic Potential	0.62	0.27, 0.96	.001
Strong Leader	−0.10	−0.45, 0.25	.602
Bribery Game	−0.44	0.39, 0.49	< .001
Period	−0.01	−0.01, 0.00	.144
Version	0.42	0.09, 0.73	.013
Subjects	0.03	−0.15, 0.20	.777
Order	0.17	0.14, 0.21	< .001
Age	0.00	−0.02, 0.01	.852
Male	0.23	0.09, 0.37	.002
(Intercept)	−0.85	−2.23, −0.24	.109
Obs.		2716	
N		248	
Group Num.		56	
R^2		0.76	

Table S4. Multilevel model regressing z-score of contribution to calculate standardized coefficients, with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.76$.

Causes of Corruption

Here we test the predictors of player contributions, bribes, and leader behavior:

Contributions

We predict a negative interaction between game (IPGG vs BG) and leader power (S) and between game and economic potential (M). That is, stronger leaders will increase contributions in the IPGG, but decrease contributions in BG. And higher economic potential will increase contributions in the IPGG, but will have no effect or a smaller effect in the BG.

Bayesian

Model

```
zContribution ~ factor(MPCR) * factor(Cond) + factor(LeaderPower) *
factor(Cond) + factor(Cond) + Period + factor(Version) + Subjects +
as.numeric(Order) + scale(age) + male
```

Results

	Coefficient	95% CI	p-value
High Economic Potential	2.01	1.39, 2.65	$< .001$
Bribery Game	−0.85	−1.16, −0.51	$< .001$
Strong Leader	−0.12	−0.80, 0.53	.756
Period	−0.06	−0.04, 0.01	.180
Version	1.01	0.43, 1.64	.002
Subjects	0.06	−0.23, 0.34	.654
Order	0.51	0.40, 0.63	$< .001$
Age	0.01	−0.25, 0.30	.974
Male	0.97	0.44, 1.58	$< .001$
High Economic Potential: Bribery Game	−0.40	−0.71, −0.05	.006
Bribery Game: Strong Leader	−0.57	−0.94, −0.22	$< .001$
(Intercept)	−2.52	−3.95, −0.57	.002
Obs.		2716	
N		248	
Group Num.		56	
DIC		10639.34	

Table S5. MCMC GLMM regression on raw, unstandardized contribution.

	Coefficient	95% CI	p-value
High Economic Potential	0.63	0.45, 0.85	< .001
Bribery Game	−0.27	−0.37, −0.16	< .001
Strong Leader	−0.04	−0.23, 0.17	.764
Period	−0.01	−0.01, 0.00	.164
Version	0.32	0.12, 0.52	.004
Subjects	0.02	−0.07, 0.11	.668
Order	0.16	0.13, 0.20	< .001
Age	0.00	−0.08, 0.09	.952
Male	0.31	0.10, 0.48	.006
High Economic Potential: Bribery Game	−0.12	−0.23, −0.03	.018
Bribery Game: Strong Leader	−0.18	−0.29, −0.06	.002
(Intercept)	−0.80	−1.32, −0.27	.004
Obs.		2716	
N		248	
Group Num.		56	
DIC		4382.554	

Table S6. MCMC GLMM regression on z-score of contribution to calculate standardized coefficients.

We can graph these effects:

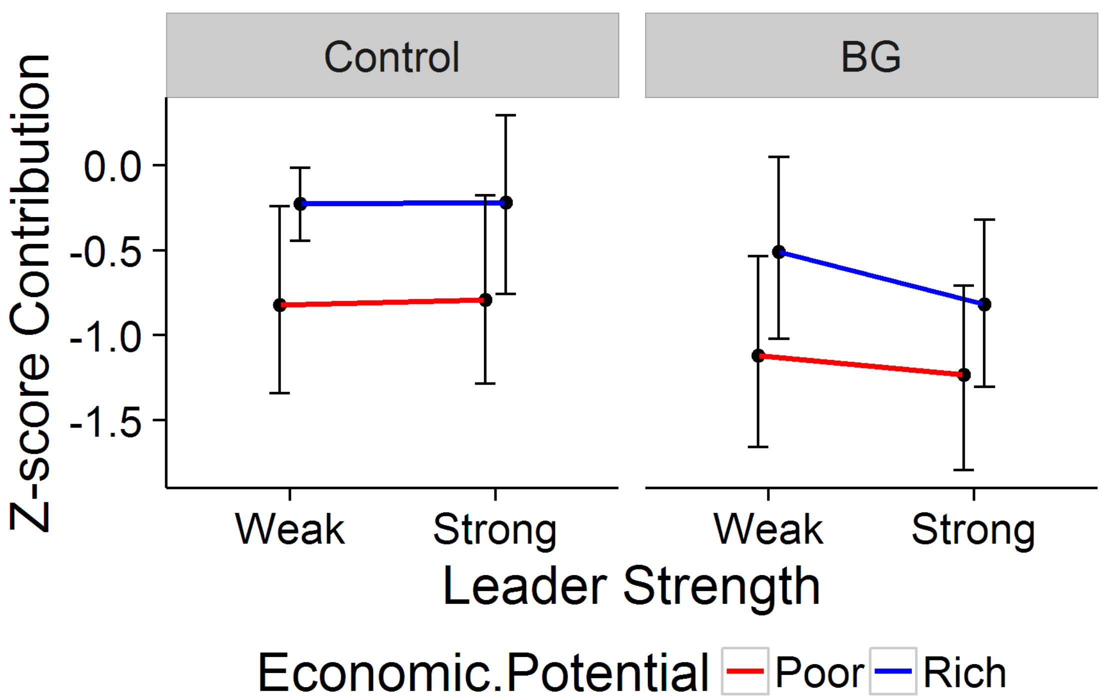


Figure S16. Comparison of contributions in the IPGG (Control) and BG for weak vs strong leaders by poor vs rich economic potential. Overall contributions are lower in BG in all

contexts. Overall contributions are higher in richer economic potential contexts. As predicted, when leaders are stronger, we see a slight increase in contributions in the IPGG, but a decrease in the BG. Also, as predicted, the effect of economic potential on increasing contributions is weaker in the BG compared to the IPGG.

Summary

These results partially support our hypothesis. Stronger leaders barely increase contributions in the IPGG, but clearly decrease contributions in the BG (as predicted). Moreover, the effect of richer economic potential is lower in the BG compared to the IPGG (as predicted).

Next, we test our prediction that stronger leaders increase bribes (rather than contributions) in the BG.

Frequentist

Model

```
cContribution ~ factor(MPCR) * factor(Cond) + factor(LeaderPower) *
factor(Cond) + factor(Cond) + Period + factor(Version) +
    Subjects + as.numeric(Order) + scale(age) + male + (1 | PID) +
(1 | GroupNum)
   Data: dat[dat$Cond == "BG" | dat$Cond == "Control", ]
```

Results

	Coefficient	95% CI	p-value
High Economic Potential	2.11	1.00, 3.22	< .001
Bribery Game	−0.86	−1.18, −0.54	< .001
Strong Leader	0.00	−1.13, 1.13	1.00
Period	−0.02	−0.04, 0.01	.149
Version	1.32	0.28, 2.32	.014
Subjects	0.06	−0.50, 0.62	.838
Order	0.47	0.35, 0.59	< .001
Age	−0.02	−0.22, 0.18	.825
Male	0.71	0.27, 1.16	.002
High Economic Potential: Bribery Game	−0.35	−0.68, −0.02	.041
Bribery Game: Strong Leader	−0.61	0.97, −0.24	.001
(Intercept)	−2.58	−5.61, 0.45	.112
Obs.		2716	
N		248	
Group Num.		56	
R^2		0.76	

Table S7. Multilevel model regressing raw, unstandardized contribution with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.76$.

	Coefficient	95% CI	p-value
High Economic Potential	0.67	0.31, 1.02	.001
Bribery Game	−0.27	−0.37, −0.17	< .001
Strong Leader	0.00	−0.36, 0.36	.999
Period	−0.01	−0.01, 0.00	.149
Version	0.42	0.09, 0.73	.014
Subjects	0.02	−0.16, 0.20	.838
Order	0.15	0.11, 0.19	< .001
Age	−0.01	−0.07, 0.06	.825
Male	0.22	0.09, 0.37	.002
High Economic Potential: Bribery Game	−0.11	−0.22, −0.01	.041
Bribery Game: Strong Leader	−0.19	−0.31, −0.08	.001
(Intercept)	−0.81	−1.77, 0.14	.112
Obs.		2716	
N		248	
Group Num.		56	
R^2		0.76	

Table S8. Multilevel model regressing z-score of contribution to calculate standardized coefficients, with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $\boldsymbol{R^2 = 0.76}$.

Bribes

We predict a positive effect of leader power (S) on bribes, but no effect or a negative effect of economic potential (M). That is, stronger leaders will increase bribes.

Bayesian

Model

```
mcmcmodel <- MCMCglmm(zBribe ~ factor(MPCR) + factor(LeaderPower) +
                        Period + factor(Version) + Subjects +
                        as.numeric(Order) + scale(age) + male,
                      random=~PID:GroupNum,
                      data=dat[dat$Cond=="BG",])
```

Results

	Coefficient	95% CI	p-value
High Economic Potential	0.28	−0.03, 0.55	.056
Strong Leader	0.57	0.25, 0.87	< .001
Period	0.02	0.01, 0.03	< .001
Version	−0.34	−0.66, −0.04	.034
Subjects	−0.05	−0.20, 0.07	.414
Order	−0.14	−0.27, −0.01	.030
Age	−0.08	−0.22, 0.05	.246
Male	−0.05	−0.34, 0.26	.776
(Intercept)	0.52	−0.38, 1.38	.240
Obs.		1420	
N		176	
Group Num.		45	
DIC		3291.46	

Table S9. MCMC GLMM regression on raw, unstandardized bribe.

	Coefficient	95% CI	p-value
High Economic Potential	0.22	0.00, 0.44	.052
Strong Leader	0.45	0.19, 0.67	< .001
Period	0.02	0.01, 0.03	.002
Version	−0.28	−0.52, −0.04	.020
Subjects	−0.04	−0.14, 0.06	.448
Order	−0.11	−0.22, −0.01	.040
Age	−0.06	−0.17, 0.05	.272
Male	−0.04	−0.27, 0.19	.756
(Intercept)	0.43	−0.33, 1.06	.238
Obs.		1420	
N		176	
Group Num.		45	
DIC		2593.12	

Table S10. MCMC GLMM regression on z-score of bribe to calculate standardized coefficients.

We can graph these effects:

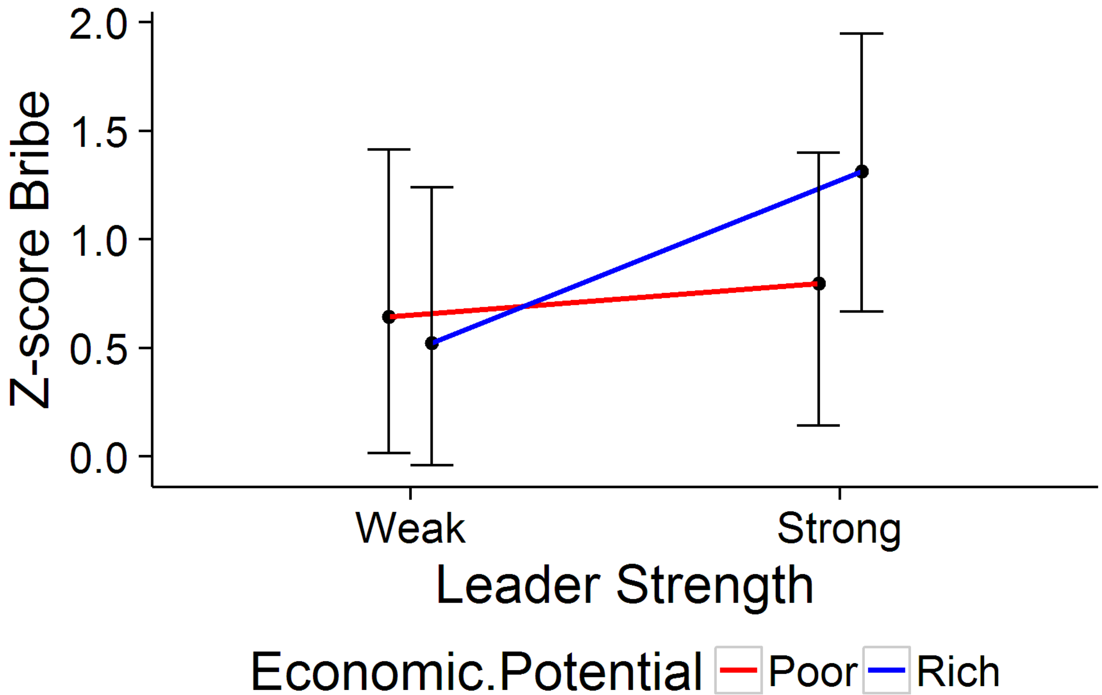


Figure S17. Comparison of bribes in the BG for weak vs strong leaders by poor vs rich economic potential. As predicted, when leaders are stronger, we see an increase in bribes.

Summary

As predicted, we find that stronger leaders extract larger bribes. Surprisingly, we find some possible evidence that this effect is stronger in richer economic potential than poorer. If these results generalize, one possible explanation for this is that Leader's and players have a non-zero norm for prosocial contributions. Leader's use punishment to achieve this minimum contribution. Since this contribution is more likely to be met in a richer economic potential context, leaders use more of their punitive power to extract bribes.

Next, we look at what predicts when Leaders will punish. If this hypothesis about leader's expecting a minimum contribution to the pubic good is correct, then we should see contributions predict punishment in the BG (not just bribes).

Frequentist

Model

```
model <- lmer(cBribe ~ factor(MPCR) + factor(LeaderPower) +
                        Period + factor(Version) + Subjects +
                        as.numeric(Order) + scale(age) + male +
                        (1|PID) + (1|GroupNum),
```

```
data=dat[dat$Cond=="BG",])
```

Results

	Coefficient	95% CI	p-value
High Economic Potential	0.36	−0.13, 0.85	.172
Strong Leader	0.74	0.24, 1.23	.009
Period	0.02	0.01, 0.03	.001
Version	−0.76	−1.26, −0.21	.007
Subjects	−0.04	−0.27, 0.19	.765
Order	−0.13	−0.36, 0.09	.268
Age	0.01	−0.11, 0.12	.873
Male	−0.11	−0.35, 0.14	.380
(Intercept)	0.56	−0.88, 2.00	.475
Obs.		1420	
N		176	
Group Num.		45	
R^2		0.70	

Table S11. Multilevel model regressing raw, unstandardized bribe with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.70$.

	Coefficient	95% CI	p-value
High Economic Potential	0.29	−0.10, 0.66	.172
Strong Leader	0.58	0.19, 0.96	.009
Period	0.02	0.01, 0.03	.001
Version	−0.59	−0.99, −0.16	.007
Subjects	−0.03	−0.21, 0.15	.765
Order	−0.11	−0.28, 0.07	.268
Age	0.01	−0.09, 0.09	.873
Male	−0.09	−0.28, 0.11	.380
(Intercept)	0.44	−0.69, 1.56	.475
Obs.		1420	
N		176	
Group Num.		45	
R^2		0.70	

Table S12. Multilevel model regressing z-score of contribution to calculate standardized coefficients, with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.70$.

Punishment

We predict that more punishments will be allocated to bribes and that Leader's will be less tolerant of small bribes when they have more power (they'll punish small bribes more).

Bayesian

Model

```
model <- lmer(cBribe ~ cPunishment ~ factor(MPCR) * cBribe +
factor(LeaderPower) * cBribe + factor(MPCR) * cContribution +
factor(LeaderPower) * cContribution + Period + factor(Version) +
Subjects + as.numeric(Order) + scale(age) + male,
                      data=dat[dat$Cond=="BG",])
```

Results

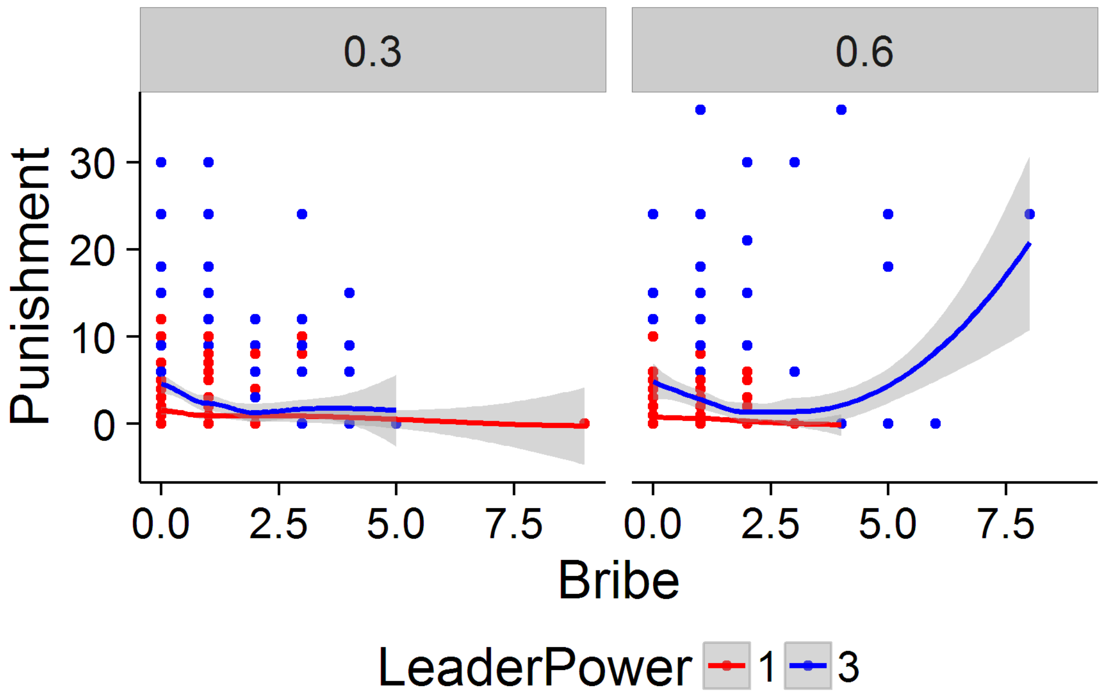
	Coefficient	95% CI	p-value
High Economic Potential	0.49	−0.26, 1.24	.226
Bribery	−0.85	−1.28, −0.42	< .001
Strong Leader	1.79	1.04, 2.54	< .001
Contribution	−0.22	−0.39, −0.06	.008
Period	0.07	−0.01, 0.13	.072
Version	−0.44	−1.22, 0.30	.244
Subjects	0.23	−0.06, 0.57	.150
Order	−0.44	−0.78, −0.09	.022
Age	−0.13	−0.51, 0.27	.462
Male	0.45	−0.34, 1.18	.250
High Economic Potential: Bribery	0.55	0.07, 1.04	.028
Bribery: Strong Leader	−0.47	−0.99, 0.03	.074
High Economic Potential: Contribution	0.23	0.01, 0.47	.056
Strong Leader: Contribution	−0.35	−0.54, −0.15	.004
(Intercept)	−1.63	−3.82, 0.61	.156
Obs.		1420	
N		176	
Group Num.		45	
DIC		8032.33	

Table S13. MCMC GLMM regression on raw, unstandardized punishment.

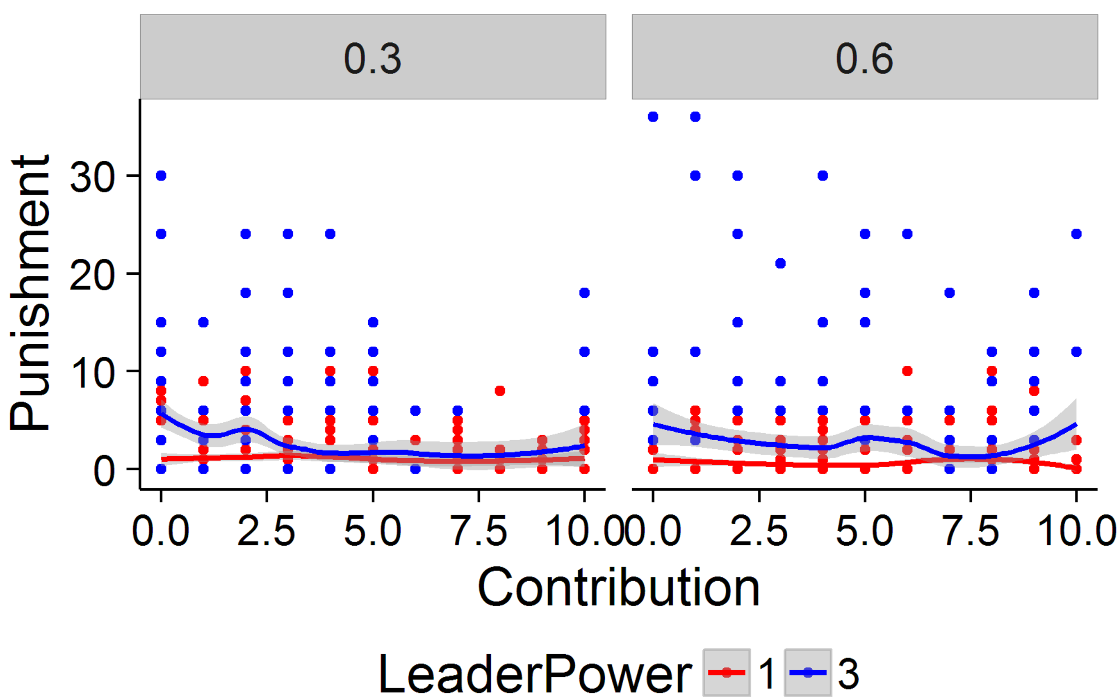
	Coefficient	95% CI	p-value
High Economic Potential	0.11	−0.07, 0.29	.232
Bribery	−0.25	−0.37, −0.12	< .001
Strong Leader	0.41	0.24, 0.58	< .001
Contribution	−0.16	−0.28, −0.02	.010
Period	0.02	0.00, 0.03	.070
Version	−0.10	−0.28, 0.07	.280
Subjects	0.05	−0.02, 0.13	.140
Order	−0.10	−0.18, −0.02	.022
Age	−0.03	−0.12, 0.05	.466
Male	0.10	−0.07, 0.27	.248
High Economic Potential: Bribery	0.16	0.03, 0.31	.024
Bribery: Strong Leader	−0.14	−0.28, 0.01	.070
High Economic Potential: Contribution	0.17	−0.01, 0.32	.040
Strong Leader: Contribution	−0.25	−0.40, −0.11	< .001
(Intercept)	−0.39	−0.90, 0.10	.140
Obs.		1420	
N		176	
Group Num.		45	
DIC		3801.53	

Table S14. MCMC GLMM regression on z-score of punishment to calculate standardized coefficients.

We can graph contributions and bribes against punishment (this is the actual data, not controlling for effects):


0.3
0.6
Punishment
30
20
10
0
0.0
2.5
5.0
7.5
Bribe
LeaderPower
1
3

(a)


0.3
0.6
Punishment
30
20
10
0
0.0
2.5
5.0
7.5
10.0
Contribution
LeaderPower
1
3

(b)

Figure S18. Punishment plotted against (a) Bribes and (b) Contribution. Loess curve to show pattern. High contributions, and certainly high bribes, are rare, but the overall pattern suggests more punishments allocated for both low contributions and low bribes.

Summary

As predicted, we find that more powerful leaders are more punitive towards smaller bribes (though this effect is marginally significant). Surprisingly, even in the BG, more powerful leaders are also more likely to punish smaller contributions. These results suggest that Leader's possess either a pro-social or anti-corruption norm. Curiously, smaller bribes and contributions both receive smaller punishments when in a richer economic context. It is possible that in this richer economic context, contributions are more in line with Leader expectations, based purely on the norm and economic potential. If this is the case, we should expect that Leaders are more likely to do nothing or to accept bribes when economic potential is greater.

Frequentist

Model

```
model <- lmer(zPunishment ~ factor(MPCR)*zBribe +
factor(LeaderPower)*zBribe + factor(MPCR)*zContribution +
factor(LeaderPower)*zContribution +
                Period + factor(Version) + Subjects +
                as.numeric(Order) + scale(age) + male +
                (1|PID) + (1|GroupNum),
              data=dat[dat$Cond=="BG",])
```

Results

	Coefficient	95% CI	p-value
High Economic Potential	0.70	−0.35, 1.65	.214
Bribery	−0.93	−1.34, −0.47	< .001
Strong Leader	2.04	0.99, 3.00	.001
Contribution	−0.27	−0.45, −0.07	.008
Period	0.07	−0.01, 0.14	.068
Version	−0.93	−1.92, 0.22	.112
Subjects	0.23	−0.21, 0.67	.363
Order	−0.40	−0.85, 0.05	.128
Age	−0.02	−0.39, 0.30	.919
Male	0.40	−0.33, 1.13	.303
High Economic Potential: Bribery	0.66	0.12, 1.16	.014
Bribery: Strong Leader	−0.65	−1.13, −0.06	.02
High Economic Potential: Contribution	0.23	0.00, 0.47	.062
Strong Leader: Contribution	−0.38	−0.59, −0.15	.001
(Intercept)	−1.67	−4.55, 1.19	.307
Obs.		1420	
N		176	
Group Num.		45	
R^2		0.34	

Table S15. Multilevel model regressing raw, unstandardized punishment with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $\boldsymbol{R^2 = 0.76}$.

	Coefficient	95% CI	p-value
High Economic Potential	0.16	−0.08, 0.37	.214
Bribery	−0.27	−0.39, −0.14	< .001
Strong Leader	0.46	0.22, 0.68	.001
Contribution	−0.19	−0.32, −0.05	.008
Period	0.02	0.00, 0.03	.068
Version	−0.21	−0.43, 0.05	.112
Subjects	0.05	−0.05, 0.15	.363
Order	−0.09	−0.19, 0.01	.128
Age	0.00	−0.09, 0.07	.919
Male	0.09	−0.08, 0.25	.303
High Economic Potential: Bribery	0.19	0.04, 0.33	.014
Bribery: Strong Leader	−0.19	−0.33, −0.02	.016
High Economic Potential: Contribution	0.17	0.00, 0.34	.062
Strong Leader: Contribution	−0.27	−0.42, −0.10	.001
(Intercept)	−0.38	−1.03, 0.27	.307
Obs.		1420	
N		176	
Group Num.		45	
R^2		0.34	

Table S16. Multilevel model regressing raw, unstandardized punishment with random effects for participants within groups. The variance explained by both fixed and random factors[34,35] is $R^2 = 0.34$.

Leader Decisions

We predict that stronger leaders should accept more bribes.

Bayesian

Model

```
mcmcmodel.bribe <- MCMCglmm(LeaderDec_Bribe ~
factor(MPCR)+factor(LeaderPower) + Period + factor(Version) + Subjects
+ as.numeric(Order) + scale(age) + male,
                      random=~PID:GroupNum, data=dat[dat$Cond=="BG",],
family = "categorical", burnin=50000,nitt=1000000,thin=5000)
mcmcmodel.punish <- MCMCglmm(LeaderDec_Punish ~
factor(MPCR)+factor(LeaderPower) + Period + factor(Version) + Subjects
+ as.numeric(Order) + scale(age) + male,
                      random=~PID:GroupNum, data=dat[dat$Cond=="BG",],
family = "categorical", burnin=50000,nitt=1000000,thin=5000)
mcmcmodel.nothing <- MCMCglmm(LeaderDec_Nothing ~
factor(MPCR)+factor(LeaderPower) + Period + factor(Version) + Subjects
+ as.numeric(Order) + scale(age) + male,
                      random=~PID:GroupNum, data=dat[dat$Cond=="BG",],
family = "categorical", burnin=50000,nitt=1000000,thin=5000)
mcmcmodel.punish.pgg <- MCMCglmm(LeaderDec_Punish ~ factor(MPCR) +
factor(LeaderPower) + Period + factor(Version) + Subjects +
as.numeric(Order) + scale(age) + male,
```

```
                                random=~PID:GroupNum,
data=levi_subset_var[levi_subset_var$Cond=="Control",], family =
"categorical", burnin=50000,nitt=1000000,thin=5000)
```

Results

	Accept Bribe	Punish	Do Nothing
High Economic Potential	1.33 [0.79,1.98]	0.81 [0.50,1.19]	0.84 [0.38,1.49]
Strong Leader	1.91 [1.11,3.03]	1.06 [0.69,1.69]	0.38 [0.14,0.67]
Period	1.02 [0.98,1.09]	1.00 [0.95,1.05]	0.97 [0.91,1.03]
Version	0.92 [0.48,1.46]	1.29 [0.76,1.91]	1.10 [0.37,1.87]
Subjects	1.04 [0.81,1.26]	1.19 [0.98,1.45]	0.82 [0.58,1.08]
Order	1.21 [0.94,1.49]	0.81 [0.64,0.98]	0.93 [0.63,1.24]
Age	0.91 [0.77,1.08]	0.85 [0.69,1.03]	1.24 [1.02,1.52]
Male	0.68 [0.49,0.93]	1.32 [0.93,1.71]	1.33 [0.84,1.80]
(Intercept)	0.74 [0.05,1.91]	0.16 [0.01,0.37]	1.94 [0.05,6.79]
Obs.	1396	1396	1396
N	175	175	175
Groups	45	45	45
DIC	37.93	20.07	20.47

Table S17. MCMC GLMM categorical regression (equivalent to logistic regression) of each leader decision against the other two decisions.

	Do Nothing
High Economic Potential	0.84 [0.51,1.19]
Strong Leader	1.06 [0.63,1.37]
Period	1.03 [0.98,1.07]
Version	1.33 [0.85,1.95]
Subjects	1.14 [0.96,1.43]
Order	1.40 [1.17,1.62]
Age	0.95 [0.78,1.21]
Male	1.17 [0.78,1.53]
(Intercept)	0.40 [0.07,0.85]
Obs.	1296
N	161
Groups	43
DIC	35.08

Table S18. MCMC GLMM categorical regression (equivalent to logistic regression) of doing nothing compared to punishing in Control (IPGG).

Summary

In line with our predictions, the only robust effect is that more powerful leaders are almost twice as likely to accept bribes and more than 2.5 times less likely to do nothing.

Yet, since leaders are also punishing for low contributions, our results suggest that something other than pure rational behavior as captured by our model is at play. Cultural evolutionary models suggest that people may internalize norms, which then influence their behavior. Next we test the effect of exposure to norms on player and Leader behaviors.

Frequentist

Model

```
model.bribe <- multinom(LeaderDec_Bribe ~
factor(MPCR)+factor(LeaderPower) +
                          Period + factor(Version) + Subjects +
                          as.numeric(Order) + scale(age) + male,
                          random=~ 1|LeaderID/GroupNum,
                       data=dat[dat$Cond=="BG",], family = binomial)
```

Results

	Accept Bribe	Punish	Do Nothing
High Economic Potential	1.43 [1.14,1.78]	0.75 [0.57,0.97]	0.80 [0.61,1.05]
Strong Leader	1.57 [1.25,1.98]	1.05 [0.81,1.38]	0.49 [0.37,0.65]
Period	1.01 [0.97,1.05]	1.01 [0.97,1.05]	0.98 [0.94,1.02]
Version	0.75 [0.59,0.95]	1.38 [1.04,1.83]	1.09 [0.81,1.45]
Subjects	1.02 [0.93,1.12]	1.19 [1.06,1.33]	0.80 [0.71,0.90]
Order	1.14 [1.03,1.25]	0.88 [0.78,0.98]	0.95 [0.84,1.07]
Age	0.87 [0.78,0.98]	0.84 [0.70,1.00]	1.29 [1.13,1.46]
Male	0.67 [0.53,0.84]	1.30 [1.00,1.69]	1.35 [1.02,1.78]
(Intercept)	0.73 [0.37,1.45]	0.13 [0.06,0.29]	1.44 [0.62,3.36]
Obs.	1396	1396	1396
N	175	175	175
Groups	45	45	45
AIC	1915.58	1500.68	1434.64

Table S19. Multilevel logistic regression of each leader decision against the other two decisions, with random effects for players within groups.

	Do Nothing
High Economic Potential	0.82 [0.64,1.04]
Strong Leader	0.99 [0.77,1.26]
Period	1.03 [0.99,1.07]
Version	1.27 [1.00,1.62]
Subjects	1.13 [0.99,1.29]
Order	1.36 [1.22,1.52]
Age	0.96 [0.85,1.08]
Male	1.15 [0.90,1.46]
(Intercept)	0.34 [0.15,0.79]
Obs.	1296
N	161
Groups	43
AIC	1667.04

Table S20. Multilevel logistic regression of each leader decision of doing nothing compared to punishing in Control (IPGG).

Exposure to Norms

Here we test how exposure to corruption norms affect behavior in our game. We do so by using our exposure score (a mean of the corruption perceptions of the countries the participant has lived in) and the heritage corruption score (a mean of the corruption perceptions of the countries the participant has an ethnic heritage). Since there is no incentive to offer bribes or contribute, except when compelled to do so by punishment, we predict that exposure to norms should primarily affect Leader decisions. Nonetheless, internalized norms may also affect the behavior of players in contributing and bribing.

We want to test the effect of direct exposure to corruption norms, but we would also like to control for heritage exposure (i.e. do these norms affect individuals who have lived in these countries, but are not natives to these corrupt countries). Similarly, we want to see the effect of heritage norms, but also look at the effect on second generation migrants and beyond, by controlling for actual direct exposure by having lived in a more corrupt country. The correlation between the direct exposure and heritage measures of corruption is $r = 0.67, p < .001$. To check if we can include both variables in our model, we check the Variance Inflation Factor on a fixed effect version of our model. These are reported for all models below.

We are interested in the effect norms have on player behavior as well as leader behavior. In each case, we run a model with player norms, with leader norms, and with both player and leader norms.

Summary

All the analyses tell a consistent story—the participants in our experiment whose families came from countries with high corruption, were themselves less likely to engage in corruption. We see no effect of direct exposure to corruption, until we control for these individuals. Then we see that direct exposure to corruption norms results in increased corrupt behavior—i.e. in our Canadian sample, those who have lived in corrupt countries from which they do not derive their heritage behave in more corrupt ways. These data are consistent with second generation migrants acculturating to local Canadian norms and also with selection in the previous generation for low corruption—i..e. those who preferred less corruption moved to Canada in the previous generation. Our data do not allow us to distinguish between these explanations, however, assuming no differential selection pressures between generations, the behavior of Canadians with direct exposure to corruption norms suggests this might be a case of acculturation (that is those with direct exposure behave in a more corrupt mannner, suggesting that the parents of those with a heritage that included corrupt nations were also more corrupt, but their children are less corrupt).

Contributions

Do corruption norms affect contributions? We look at the effect of corruption norms in the BG.

VIF

```
model <- lm(zContribution ~ factor(MPCR)+
              factor(LeaderPower)+
              zPlayerExposureCorruption +
              zPlayerHeritageCorruption +
              zLeaderExposureCorruption +
              zLeaderHeritageCorruption +
              Period + factor(Version) + Subjects +
              as.numeric(Order) + scale(age) + male,
            data=dat.norms[dat.norms$Cond=="BG",])
```

All corruption norm variables have VIF<2.5.

Variable	VIF
Player Exposure Corruption Score	1.83
Player Heritage Corruption Score	1.87
Leader Exposure Corruption Score	1.62
Leader Heritage Corruption Score	1.68

Table S21. VIF Scores for OLS regression on contribution.

Bayesian

Effect of Norms on Contributions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.54***	0.53***	0.52***	0.54***	0.54***	0.53***	0.53***
Strong Leader	−0.18	−0.20	−0.19	−0.18	−0.18	−0.18	−0.20
Player Exposure Corruption Score	0.00		−0.06				−0.05
Player Heritage Corruption Score		0.06	0.09				0.08
Leader Exposure Corruption Score				0.01		0.01	0.01
Leader Heritage Corruption Score					0.01	0.01	0.01
Period	−0.02**	−0.02***	−0.02***	−0.02***	−0.02**	−0.02***	−0.02***
Version	0.31*	0.31*	0.31*	0.31*	0.32**	0.31*	0.31*
Subjects	0.00	0.00	−0.01	0.00	0.00	0.00	−0.01
Order	0.18***	0.19***	0.19***	0.18***	0.19***	0.19***	0.19***
Age	0.09	0.08	0.08	0.09	0.09	0.09	0.08
Male	0.29*	0.29*	0.29*	0.29*	0.28*	0.28*	0.28*
(Intercept)	−0.98**	−0.98**	−0.94**	−0.98**	−0.99**	−0.97**	−0.94**
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
DIC	2140.53	2141.27	2140.08	2143.13	2141.65	2144.77	2144.90

Table S22. MCMC GLMM regression on z-score of contribution.

We find no evidence that leader corruption norms affect contributions. We find a small effect suggesting that players with a heritage that includes countries with high corruption norms actually contribute more and players with direct exposure to corruption contribute less, but this effect is not significant. Note that leadership is randomly assigned, so the effect of leaders must occur via shaping the norms of the groups they are in. We can test this by checking if mean contributions are higher in groups where heritage corruption scores are higher.

How do corruption norms in groups affect mean of contributions?

We calculate the mean contribution within each group and predict this with the mean of corruption norms.

	Coefficient	95% CI	p-value
High Economic Potential	0.58	0.15, 1.02	.006
Strong Leader	−0.07	−0.54, 0.35	.744
Exposure Corruption Score	−0.06	−0.73, 0.53	.866
Heritage Corruption Score	0.01	−0.59, 0.68	.968
Version	0.23	−0.19, 0.71	.340
Subjects	−0.04	−0.23, 0.18	.678
Order	0.16	−0.05, 0.35	.118
Mean Age	0.11	−0.11, 0.32	.318
Percent Male	0.36	−0.49, 1.21	.402
(Intercept)	−0.87	−2.18, 0.61	.208
Obs.		1396	
N		175	
Group Num.		45	
DIC		112.30	

Table S23. MCMC GLMM regression of mean of z-score of contributions in each group on mean corruption scores of players in the group.

No clear patterns emerge at the group level. Next we look at how corruption norms affect bribing behavior.

Frequentist

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.59**	0.59**	0.58**	0.59**	0.59**	0.59**	0.58**
Strong Leader	−0.13	−0.15	−0.15	−0.13	−0.14	−0.14	−0.15
Player Exposure Corruption Score	0.00		−0.06				−0.06
Player Heritage Corruption Score		0.06	0.10				0.10
Leader Exposure Corruption Score				0.01		0.00	0.00
Leader Heritage Corruption Score					0.01	0.01	0.01
Period	−0.02***	−0.02***	−0.02***	−0.02***	−0.02***	−0.02***	−0.02***
Version	0.26	0.25	0.27	0.26	0.26	0.26	0.27
Subjects	−0.01	−0.02	−0.02	−0.01	−0.01	−0.01	−0.02
Order	0.16	0.15	0.15	0.16	0.15	0.15	0.15
Age	0.06	0.06	0.05	0.06	0.06	0.06	0.05
Male	0.24**	0.24**	0.24**	0.24**	0.24**	0.24**	0.24**
(Intercept)	−0.84	−0.82	−0.81	−0.84	−0.83	−0.83	−0.79
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
R^2	0.75	0.75	0.75	0.75	0.75	0.75	0.75

Table S24. Multilevel model regressing z-score of contribution.

	Coefficient	95% CI	p-value
High Economic Potential	0.59	0.14, 1.04	.011
Strong Leader	−0.08	−0.55, 0.40	.742
Exposure Corruption Score	−0.06	−0.67, 0.55	.840
Heritage Corruption Score	0.01	−0.62, 0.64	.968
Version	0.22	−0.26, 0.70	.353
Subjects	−0.03	−0.25, 0.18	.748
Order	0.15	−0.06, 0.36	.156
Mean Age	0.11	−0.12, 0.34	.347
Percent Male	0.34	−0.50, 1.18	.423
(Intercept)	−0.88	−2.31, 0.55	.220
Obs.		1396	
N		175	
Group Num.		45	
R^2		0.29	

Table S25. OLS regression of mean of z-score of contributions in each group on mean corruption scores of players in the group.

Bribes

VIF

Variable	VIF
Player Exposure Corruption Score	1.22
Player Heritage Corruption Score	1.83
Leader Exposure Corruption Score	1.62
Leader Heritage Corruption Score	1.68

Table S26. VIF Scores for OLS regression on bribes.

Effect of Norms on Bribes

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.21	0.22	0.23	0.22	0.23	0.23	0.24*
Strong Leader	0.44***	0.46***	0.46***	0.44***	0.45***	0.45***	0.48***
Player Exposure Corruption Score	−0.03		0.04				0.04
Player Heritage Corruption Score		−0.08	−0.10				−0.10
Leader Exposure Corruption Score				−0.01		0.01	0.01
Leader Heritage Corruption Score					−0.03	−0.04	−0.04
Period	0.02**	0.02***	0.02***	0.02**	0.02**	0.02**	0.02**
Version	−0.28*	−0.28*	−0.28*	−0.28*	−0.29*	−0.29*	−0.28*
Subjects	−0.05	−0.04	−0.04	−0.04	−0.03	−0.04	−0.03
Order	−0.11	−0.11*	−0.11*	−0.11*	−0.11	−0.11	−0.11*
Age	−0.06	−0.05	−0.05	−0.06	−0.06	−0.06	−0.05
Male	−0.04	−0.04	−0.04	−0.04	−0.04	−0.04	−0.05
(Intercept)	0.43	0.42	0.39	0.41	0.38	0.38	0.35
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
DIC	2559.57	2558.78	2559.65	2561.71	2558.88	2559.87	2560.08

Table S27. MCMC GLMM regression on z-score of bribe.

How do corruption norms in group affect bribe behavior in group

	Coefficient	95% CI		p-value
High Economic Potential	0.33	−0.07,	0.74	.132
Strong Leader	0.47	0.02,	0.89	.044
Exposure Corruption Score	−0.02	−0.58,	0.56	.936
Heritage Corruption Score	−0.06	−0.63,	0.50	.832
Version	−0.37	−0.84,	0.06	.098
Subjects	0.01	−0.18,	0.21	.956
Order	−0.10	−0.30,	0.08	.344
Mean Age	−0.19	−0.40,	0.02	.088
Percent Male	0.03	−0.73,	0.87	.914
(Intercept)	0.20	−1.11,	1.49	.764
Obs.		1396		
N		175		
Group Num.		45		
DIC		107.33		

Table S28. MCMC GLMM regression of mean of z-score of bribes in each group on mean corruption scores of players in the group.

Again, similar to contributions and not statistically significant, we find that direct exposure results in higher bribes, but heritage lower bribes.

Frequentist

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.29	0.30	0.30	0.29	0.30	0.30	0.31
Strong Leader	0.58**	0.60**	0.60**	0.58**	0.59**	0.59**	0.61**
Player Exposure Corruption Score	−0.02		0.04				0.04
Player Heritage Corruption Score		−0.07	−0.10				−0.10
Leader Exposure Corruption Score				−0.10		0.02	0.02
Leader Heritage Corruption Score					−0.03	−0.04	−0.04
Period	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**
Version	−0.59**	−0.60**	−0.60**	−0.60**	−0.60**	−0.61**	−0.61**
Subjects	−0.03	−0.02	−0.02	−0.03	−0.02	−0.02	−0.02
Order	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10	−0.10
Age	0.01	0.02	0.02	0.01	0.01	0.01	0.02
Male	−0.08	−0.09	−0.09	−0.08	−0.08	−0.08	−0.09
(Intercept)	0.42	0.39	0.38	0.42	0.39	0.39	0.35
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
R^2	0.70	0.70	0.70	0.70	0.70	0.70	0.70

Table S29. Multilevel model regressing z-score of bribe.

	Coefficient	95% CI	p-value
High Economic Potential	0.33	−0.09, 0.76	.119
Strong Leader	0.47	0.02, 0.92	.043
Exposure Corruption Score	−0.04	−0.62, 0.53	.883
Heritage Corruption Score	−0.04	−0.64, 0.56	.898
Version	−0.36	−0.81, 0.09	.116
Subjects	0.00	−0.20, 0.21	.970
Order	−0.09	−0.29, 0.11	.378
Mean Age	−0.19	−0.42, 0.03	.084
Percent Male	0.05	−0.75, 0.84	.906
(Intercept)	0.18	−1.18, 1.53	.795
Obs.		1396	
N		175	
Group Num.		45	
R^2		0.27	

Table S30. OLS regression of mean of z-score of bribes in each group on mean corruption scores of players in the group.

Leader Decision

Players have no incentive to offer bribes, other than to avoid punishment. If exposure to norms affect bribery, we should expect that leader's who have been directly exposed to more corrupt norms accept more bribes (rather than punishing or doing nothing).

VIF

Variable	VIF
Player Exposure Corruption Score	1.84
Player Heritage Corruption Score	1.89
Leader Exposure Corruption Score	1.64
Leader Heritage Corruption Score	1.72

Table S31. VIF Scores for logistic regression on leader decision to accept bribe compared to not accepting bribe.

Variable	VIF
Player Exposure Corruption Score	1.81
Player Heritage Corruption Score	1.83
Leader Exposure Corruption Score	1.67
Leader Heritage Corruption Score	1.70

Table S32. VIF Scores for logistic regression on leader decision to punish compared to not punishing.

Variable	VIF
Player Exposure Corruption Score	1.66
Player Heritage Corruption Score	1.69
Leader Exposure Corruption Score	1.62
Leader Heritage Corruption Score	1.71

Table S33. VIF Scores for logistic regression on leader decision to do nothing compared to not doing nothing.

Bayesian

Model

```
mcmcmodel.bribe <- MCMCglmm(LeaderDec_Bribe ~
factor(MPCR)+factor(LeaderPower) +
                              zPlayerExposureCorruption +
                              zPlayerHeritageCorruption +
                              Period + factor(Version) + Subjects +
as.numeric(Order) +
                              scale(age) + male,
                            random=~LeaderID:GroupNum,
data=dat.norms[dat.norms$Cond=="BG",], family = "categorical",
burnin=50000,nitt=1000000,thin=5000)
mcmcmodel.bribe <- MCMCglmm(LeaderDec_Punish ~
factor(MPCR)+factor(LeaderPower) +
                              zPlayerExposureCorruption +
                              zPlayerHeritageCorruption +
                              Period + factor(Version) + Subjects +
as.numeric(Order) +
                              scale(age) + male,
                            random=~LeaderID:GroupNum,
data=dat.norms[dat.norms$Cond=="BG",], family = "categorical",
burnin=50000,nitt=1000000,thin=5000)
mcmcmodel.bribe <- MCMCglmm(LeaderDec_Nothing ~
factor(MPCR)+factor(LeaderPower) +
                              zPlayerExposureCorruption +
                              zPlayerHeritageCorruption +
                              Period + factor(Version) + Subjects +
as.numeric(Order) +
                              scale(age) + male,
                            random=~LeaderID:GroupNum,
data=dat.norms[dat.norms$Cond=="BG",], family = "categorical",
burnin=50000,nitt=1000000,thin=5000)
```

Results

Accept Bribe

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	1.36 [0.66,2.11]	1.36 [0.74,2.04]	1.38 [0.72,2.07]	1.36 [0.69,2.08]	1.38 [0.68,2.09]	1.42 [0.64,2.24]	1.37 [0.65,2.21]
Strong Leader	1.98 [1.09,3.20]	2.09 [1.19,3.62]	2.18 [1.10,3.48]	2.02 [0.86,3.09]	1.98 [1.03,3.05]	2.12 [1.26,3.41]	2.14 [1.18,3.36]
Player Exposure Corruption Score	0.92 [0.80,1.06]		1.22 [1.00,1.47]				1.22 [1.01,1.44]
Player Heritage Corruption Score		0.74 [0.62,0.84]	0.65 [0.53,0.78]				0.65 [0.54,0.79]
Leader Exposure Corruption Score				0.95 [0.70,1.26]		1.17 [0.80,1.57]	1.16 [0.85,1.57]
Leader Heritage Corruption Score					0.82 [0.62,1.06]	0.76 [0.56,1.09]	0.74 [0.50,1.01]
Period	1.03 [0.98,1.08]	1.03 [0.98,1.07]	1.03 [0.99,1.09]	1.03 [0.99,1.08]	1.03 [0.98,1.07]	1.03 [0.96,1.07]	1.03 [0.97,1.08]
Version	0.94 [0.60,1.50]	0.94 [0.49,1.53]	0.91 [0.49,1.43]	0.89 [0.43,1.53]	0.87 [0.47,1.32]	0.89 [0.39,1.53]	0.92 [0.47,1.51]
Subjects	1.05 [0.82,1.29]	1.07 [0.84,1.27]	1.07 [0.79,1.31]	1.04 [0.82,1.32]	1.08 [0.85,1.35]	1.09 [0.82,1.32]	1.08 [0.85,1.35]
Order	1.22 [0.93,1.53]	1.21 [0.97,1.52]	1.20 [0.92,1.53]	1.19 [0.90,1.47]	1.23 [0.91,1.52]	1.22 [0.90,1.49]	1.20 [0.92,1.46]
Age	0.90 [0.75,1.06]	0.94 [0.77,1.10]	0.96 [0.82,1.13]	0.90 [0.74,1.07]	0.90 [0.76,1.07]	0.90 [0.75,1.08]	0.94 [0.80,1.12]
Male	0.70 [0.51,0.95]	0.69 [0.49,0.87]	0.68 [0.49,0.91]	0.71 [0.50,0.97]	0.71 [0.53,0.96]	0.71 [0.51,0.90]	0.71 [0.46,0.88]
(Intercept)	0.65 [0.04,1.56]	0.57 [0.09,1.61]	0.59 [0.05,1.47]	0.73 [0.11,2.25]	0.61 [0.04,1.73]	0.58 [0.03,1.67]	0.57 [0.05,1.54]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
DIC	36.56	36.22	35.86	36.63	36.42	35.97	36.13

Table S34. MCMC GLMM categorical regression (equivalent to logistic regression) for leader decision to accept bribe compared to not accepting bribes.

Punish

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.78 [0.52,1.17]	0.79 [0.49,1.16]	0.77 [0.45,1.13]	0.79 [0.50,1.12]	0.78 [0.47,1.18]	0.81 [0.47,1.15]	0.79 [0.41,1.14]
Strong Leader	1.15 [0.73,1.69]	1.13 [0.50,1.63]	1.09 [0.59,1.57]	1.11 [0.66,1.66]	1.10 [0.69,1.66]	1.11 [0.63,1.60]	1.08 [0.60,1.61]
Player Exposure Corruption Score	1.10 [0.94,1.28]		0.99 [0.82,1.20]				0.99 [0.81,1.19]
Player Heritage Corruption Score		1.15 [0.96,1.29]	1.17 [0.99,1.42]				1.17 [0.96,1.40]
Leader Exposure Corruption Score				1.16 [0.89,1.46]		1.16 [0.85,1.54]	1.15 [0.81,1.52]
Leader Heritage Corruption Score					1.10 [0.86,1.36]	1.04 [0.76,1.33]	1.02 [0.74,1.35]
Period	1.01 [0.96,1.06]	1.00 [0.95,1.06]	1.00 [0.95,1.06]	1.00 [0.95,1.06]	1.00 [0.95,1.07]	1.00 [0.96,1.07]	1.00 [0.95,1.06]
Version	1.24 [0.68,1.90]	1.21 [0.62,1.92]	1.24 [0.78,1.91]	1.32 [0.80,2.04]	1.28 [0.76,1.94]	1.27 [0.64,1.88]	1.26 [0.74,1.84]
Subjects	1.22 [0.98,1.46]	1.21 [0.96,1.47]	1.21 [0.92,1.50]	1.21 [0.96,1.46]	1.20 [0.96,1.44]	1.20 [0.96,1.46]	1.21 [0.95,1.47]
Order	0.82 [0.67,0.98]	0.81 [0.66,0.99]	0.82 [0.66,0.99]	0.81 [0.65,0.97]	0.82 [0.66,0.98]	0.82 [0.66,0.98]	0.81 [0.64,0.97]
Age	0.83 [0.64,1.00]	0.83 [0.69,1.03]	0.84 [0.68,1.01]	0.84 [0.67,0.99]	0.84 [0.69,1.06]	0.84 [0.67,1.00]	0.82 [0.65,0.99]
Male	1.30 [0.93,1.64]	1.31 [0.92,1.70]	1.31 [0.96,1.74]	1.32 [1.00,1.82]	1.31 [0.88,1.78]	1.34 [0.93,1.76]	1.27 [0.90,1.64]
(Intercept)	0.12 [0.01,0.31]	0.14 [0.02,0.39]	0.14 [0.02,0.38]	0.14 [0.02,0.44]	0.15 [0.01,0.37]	0.14 [0.01,0.39]	0.16 [0.02,0.39]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
DIC	19.26	19.43	19.07	19.37	19.12	18.95	18.23

Table S35. MCMC GLMM categorical regression (equivalent to logistic regression) for leader decision to punish compared to not punishing.

Do Nothing

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.92 [0.40,1.61]	0.91 [0.32,1.58]	0.86 [0.32,1.56]	0.92 [0.34,1.50]	0.84 [0.38,1.46]	0.86 [0.36,1.42]	0.81 [0.29,1.40]
Strong Leader	0.37 [0.12,0.69]	0.32 [0.15,0.60]	0.32 [0.14,0.54]	0.35 [0.13,0.65]	0.33 [0.11,0.60]	0.32 [0.08,0.62]	0.29 [0.10,0.50]
Player Exposure Corruption Score	1.02 [0.80,1.19]		0.77 [0.63,0.97]				0.79 [0.63,1.02]
Player Heritage Corruption Score		1.33 [1.14,1.59]	1.57 [1.28,1.94]				1.55 [1.25,1.89]
Leader Exposure Corruption Score				0.97 [0.67,1.39]		0.80 [0.45,1.20]	0.75 [0.42,1.09]
Leader Heritage Corruption Score					1.26 [0.74,1.80]	1.53 [0.80,2.44]	1.60 [0.99,2.43]
Period	0.96 [0.90,1.02]	0.97 [0.91,1.02]	0.97 [0.91,1.04]	0.97 [0.91,1.03]	0.96 [0.89,1.02]	0.97 [0.91,1.02]	0.97 [0.91,1.03]
Version	1.21 [0.36,2.29]	1.14 [0.45,2.17]	1.18 [0.42,2.11]	1.18 [0.51,2.27]	1.24 [0.39,2.38]	1.28 [0.43,2.43]	1.22 [0.52,2.12]
Subjects	0.81 [0.55,1.05]	0.78 [0.53,1.05]	0.78 [0.53,1.06]	0.79 [0.58,1.03]	0.79 [0.52,1.03]	0.77 [0.52,1.02]	0.73 [0.52,0.96]
Order	0.95 [0.69,1.34]	0.94 [0.65,1.26]	0.93 [0.69,1.26]	0.95 [0.70,1.31]	0.93 [0.67,1.31]	0.92 [0.60,1.24]	0.92 [0.66,1.17]
Age	1.22 [0.98,1.45]	1.19 [0.96,1.44]	1.18 [0.97,1.42]	1.22 [0.97,1.42]	1.21 [0.99,1.40]	1.22 [1.01,1.52]	1.19 [1.01,1.41]
Male	1.33 [0.86,1.77]	1.33 [0.86,1.85]	1.31 [0.86,1.76]	1.31 [0.92,1.90]	1.33 [0.90,1.74]	1.28 [0.81,1.75]	1.33 [0.86,2.01]
(Intercept)	2.03 [0.02,6.50]	2.38 [0.10,8.91]	2.43 [0.07,9.24]	1.81 [0.08,6.96]	2.38 [0.07,8.24]	2.98 [0.06,7.99]	3.01 [0.12,9.50]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
DIC	19.74	19.47	19.05	19.62	19.50	19.24	18.45

Table S36. MCMC GLMM categorical regression (equivalent to logistic regression) for leader decision to do nothing compared to not doing nothing.

Frequentist

Model

```
model <- multinom(LeaderDec_Bribe ~ factor(MPCR)+factor(LeaderPower) +
                          zPlayerExposureCorruption +
                          zPlayerHeritageCorruption +
                          Period + factor(Version) + Subjects +
                          as.numeric(Order) + scale(age) + male,
                        random=~ 1|LeaderID/GroupNum,
                        data=dat.norms[dat.norms$Cond=="BG",], family
= binomial)
model <- multinom(LeaderDec_Punish ~ factor(MPCR)+factor(LeaderPower)
+
                          zPlayerExposureCorruption +
                          zPlayerHeritageCorruption +
                          Period + factor(Version) + Subjects +
                          as.numeric(Order) + scale(age) + male,
                        random=~ 1|LeaderID/GroupNum,
```

```
                        data=dat.norms[dat.norms$Cond=="BG",], family
= binomial)
model <- multinom(LeaderDec_Nothing ~ factor(MPCR)+factor(LeaderPower)
+
                          zPlayerExposureCorruption +
                          zPlayerHeritageCorruption +
                          Period + factor(Version) + Subjects +
                          as.numeric(Order) + scale(age) + male,
                        random=~ 1|LeaderID/GroupNum,
                        data=dat.norms[dat.norms$Cond=="BG",], family
= binomial)
```

Results

Accept Bribe

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	1.42 [1.14,1.78]	1.45 [1.16,1.81]	1.47 [1.18,1.85]	1.43 [1.14,1.79]	1.48 [1.18,1.86]	1.49 [1.19,1.87]	1.54 [1.23,1.94]
Strong Leader	1.61 [1.27,2.04]	1.70 [1.34,2.16]	1.69 [1.33,2.14]	1.61 [1.28,2.04]	1.72 [1.35,2.18]	1.74 [1.37,2.20]	1.84 [1.44,2.34]
Player Exposure Corruption Score	0.98 [0.88,1.10]		1.24 [1.06,1.43]				1.22 [1.05,1.42]
Player Heritage Corruption Score		0.82 [0.73,0.91]	0.71 [0.61,0.82]				0.70 [0.61,0.82]
Leader Exposure Corruption Score				0.93 [0.83,1.04]		1.07 [0.93,1.23]	1.06 [0.92,1.22]
Leader Heritage Corruption Score					0.83 [0.74,0.93]	0.79 [0.68,0.92]	0.79 [0.68,0.92]
Period	1.01 [0.97,1.05]	1.01 [0.97,1.05]	1.01 [0.97,1.05]	1.01 [0.97,1.05]	1.01 [0.98,1.05]	1.01 [0.97,1.05]	1.01 [0.98,1.05]
Version	0.74 [0.58,0.94]	0.74 [0.58,0.94]	0.75 [0.59,0.95]	0.73 [0.58,0.93]	0.71 [0.56,0.91]	0.71 [0.56,0.91]	0.72 [0.56,0.92]
Subjects	1.02 [0.58,0.94]	1.03 [0.93,1.14]	1.04 [0.95,1.15]	1.03 [0.93,1.13]	1.05 [0.95,1.16]	1.05 [0.95,1.16]	1.07 [0.97,1.19]
Order	1.13 [1.02,1.24]	1.12 [1.02,1.24]	1.12 [1.02,1.24]	1.13 [1.02,1.25]	1.13 [1.02,1.25]	1.13 [1.02,1.25]	1.13 [1.02,1.25]
Age	0.88 [0.78,0.98]	0.90 [0.80,1.02]	0.91 [0.81,1.03]	0.88 [0.78,0.99]	0.88 [0.78,0.99]	0.88 [0.78,0.99]	0.92 [0.81,1.03]
Male	0.67 [0.53,0.85]	0.66 [0.52,0.83]	0.66 [0.52,0.83]	0.67 [0.53,0.84]	0.66 [0.52,0.83]	0.66 [0.53,0.84]	0.65 [0.51,0.82]
(Intercept)	0.72 [0.36,1.43]	0.68 [0.34,1.36]	0.64 [0.32,1.28]	0.70 [0.35,1.40]	0.61 [0.31,1.23]	0.60 [0.30,1.21]	0.53 [0.26,1.08]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
AIC	1885.43	1872.36	1866.51	1884.00	1875.63	1876.85	1859.06

Table S37. Multilevel logistic regression of each leader decision to accept bribe compared to not accepting bribe, with random effects for players within groups.

Punish

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.73 [0.55,0.95]	0.72 [0.55,0.95]	0.72 [0.55,0.94]	0.72 [0.55,0.94]	0.70 [0.53,0.92]	0.71 [0.54,0.93]	0.70 [0.54,0.92]
Strong Leader	1.10 [0.84,1.45]	1.09 [0.83,1.43]	1.09 [0.83,1.44]	1.10 [0.84,1.45]	1.05 [0.79,1.38]	1.08 [0.81,1.42]	1.06 [0.80,1.41]
Player Exposure Corruption Score	1.01 [0.89,1.15]		0.94 [0.79,1.12]				0.94 [0.79,1.12]
Player Heritage Corruption Score		1.07 [0.94,1.22]	1.11 [0.94,1.32]				1.13 [0.95,1.34]
Leader Exposure Corruption Score				1.20 [1.05,1.36]		1.14 [0.96,1.34]	1.14 [0.96,1.35]
Leader Heritage Corruption Score					1.18 [1.03,1.36]	1.09 [0.91,1.30]	1.09 [0.91,1.31]
Period	1.01 [0.97,1.06]	1.01 [0.97,1.06]	1.01 [0.97,1.06]	1.01 [0.96,1.05]	1.01 [0.97,1.06]	1.01 [0.96,1.05]	1.01 [0.96,1.05]
Version	1.32 [0.99,1.76]	1.32 [0.99,1.75]	1.31 [0.98,1.75]	1.34 [1.00,1.78]	1.35 [1.01,1.80]	1.35 [1.01,1.80]	1.34 [1.01,1.79]
Subjects	1.20 [1.08,1.35]	1.20 [1.07,1.34]	1.20 [1.07,1.34]	1.21 [1.08,1.35]	1.18 [1.06,1.33]	1.20 [1.07,1.35]	1.19 [1.06,1.34]
Order	0.86 [0.77,0.97]	0.87 [0.77,0.97]	0.87 [0.77,0.97]	0.86 [0.76,0.97]	0.86 [0.76,0.97]	0.86 [0.76,0.97]	0.86 [0.76,0.97]
Age	0.84 [0.70,1.01]	0.83 [0.69,1.00]	0.83 [0.69,1.00]	0.83 [0.69,1.00]	0.84 [0.70,1.01]	0.83 [0.69,1.00]	0.82 [0.68,0.99]
Male	1.27 [0.97,1.67]	1.28 [0.98,1.67]	1.27 [0.97,1.66]	1.32 [1.01,1.73]	1.30 [0.99,1.70]	1.32 [1.01,1.73]	1.32 [1.01,1.73]
(Intercept)	0.12 [0.06,0.28]	0.13 [0.06,0.29]	0.13 [0.06,0.29]	0.12 [0.05,0.28]	0.14 [0.06,0.32]	0.13 [0.06,0.30]	0.14 [0.06,0.31]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
AIC	1469.06	1468.02	1469.57	1462.11	1463.53	1463.24	1465.40

Table S38. Multilevel logistic regression of each leader decision to punish compared to not punishing, with random effects for players within groups.

Do Nothing

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
High Economic Potential	0.83 [0.63,1.09]	0.81 [0.62,1.07]	0.80 [0.61,1.05]	0.83 [0.63,1.09]	0.81 [0.62,1.06]	0.79 [0.60,1.04]	0.76 [0.58,1.00]
Strong Leader	0.46 [0.35,0.61]	0.43 [0.32,0.57]	0.44 [0.33,0.58]	0.46 [0.35,0.62]	0.44 [0.33,0.59]	0.43 [0.32,0.57]	0.40 [0.30,0.54]
Player Exposure Corruption Score	1.02 [0.89,1.17]		0.80 [0.68,0.95]				0.81 [0.68,0.96]
Player Heritage Corruption Score		1.27 [1.11,1.46]	1.46 [1.23,1.74]				1.47 [1.23,1.75]
Leader Exposure Corruption Score				0.94 [0.82,1.08]		0.82 [0.69,0.97]	0.82 [0.69,0.97]
Leader Heritage Corruption Score					1.12 [0.97,1.29]	1.28 [1.07,1.53]	1.29 [1.08,1.55]
Period	0.98 [0.93,1.02]	0.98 [0.93,1.02]	0.97 [0.93,1.02]	0.98 [0.94,1.02]	0.98 [0.93,1.02]	0.98 [0.94,1.02]	0.98 [0.93,1.02]
Version	1.18 [0.88,1.59]	1.18 [0.87,1.58]	1.17 [0.87,1.57]	1.18 [0.88,1.59]	1.21 [0.90,1.63]	1.23 [0.91,1.66]	1.21 [0.90,1.64]
Subjects	0.78 [0.69,0.89]	0.78 [0.69,0.88]	0.76 [0.67,0.86]	0.79 [0.70,0.89]	0.77 [0.68,0.87]	0.77 [0.68,0.87]	0.74 [0.65,0.85]
Order	0.97 [0.86,1.10]	0.98 [0.86,1.11]	0.98 [0.86,1.11]	0.97 [0.86,1.10]	0.97 [0.86,1.10]	0.97 [0.86,1.10]	0.98 [0.86,1.11]
Age	1.27 [1.12,1.45]	1.23 [1.09,1.40]	1.23 [1.08,1.40]	1.28 [1.13,1.46]	1.27 [1.12,1.45]	1.29 [1.13,1.47]	1.24 [1.08,1.41]
Male	1.36 [1.02,1.79]	1.41 [1.06,1.87]	1.41 [1.06,1.87]	1.34 [1.01,1.77]	1.37 [1.03,1.81]	1.34 [1.01,1.78]	1.40 [1.05,1.86]
(Intercept)	1.52 [0.65,3.55]	1.59 [0.68,3.73]	1.71 [0.73,4.04]	1.48 [0.63,3.46]	1.69 [0.72,4.00]	1.74 [0.74,4.11]	1.99 [0.83,4.75]
Obs.	1396	1396	1396	1396	1396	1396	1396
N	175	175	175	175	175	175	175
Groups	45	45	45	45	45	45	45
AIC	1415.05	1403.08	1398.64	1414.44	1412.65	1409.36	1394.64

Table S39. Multilevel logistic regression of each leader decision to do nothing compared to not doing nothing, with random effects for players within groups.

Next, we look at whether anti-corruption measures can return contributions to the levels seen in the IPGG, when bribery is not a option.

Cures for Corruption

Here we report the full regression discussed in the main text. Partial transparency may work by revealing or establishing contribution norms and full transparency may work by revealing contribution and bribe norms, as well as the leader's punitive preferences. Leader investment can only work by increasing a leader's tendency to punish for lack of contributions rather than lack of bribes, but this is only likely to work when economic potential is high.

Bayesian

	Coefficient	95% CI	p-value
Economic Potential	0.76	0.52, 1.00	< .001
Bribery Game	−0.21	−0.31, −0.11	< .001
Partial Transparency	−0.31	−0.39, −0.21	< .001
Full Transparency	−0.20	−0.30, −0.11	< .001
Leader Investment	−0.46	−0.55, −0.36	< .001
Strong Leader	0.17	−0.08, 0.40	.202
Period	0.01	0.00, 0.01	< .001
Version	0.24	0.05, 0.40	.006
Subjects	0.02	−0.06, 0.09	.680
Order	0.15	0.13, 0.16	< .001
Age	−0.02	−0.10, 0.05	.642
Male	0.25	0.07, 0.41	.004
Economic Potential:Bribery Game	−0.18	−0.33, −0.04	.014
Economic Potential:Partial Transparency	0.01	−0.13, 0.13	.928
Economic Potential:Full Transparency	0.05	−0.10, 0.18	.446
Economic Potential:Leader Investment	0.31	0.18, 0.44	< .001
Economic Potential:Strong Leader	−0.25	−0.59, 0.11	.172
Bribery Game:Strong Leader	−0.31	−0.46, −0.18	< .001
Part Transparency:Strong Leader	−0.22	−0.36, −0.09	< .001
Full Transparency:Strong Leader	0.15	−0.01, 0.28	.048
Leader Investment:Strong Leader	0.29	0.15, 0.44	< .001
Economic Potential:Bribery Game:Strong Leader	0.13	−0.05, 0.33	.222
Economic Potential:Partial Transparency:Strong Leader	0.09	−0.11, 0.28	.404
Economic Potential:Full Transparency:Strong Leader	−0.25	−0.46, −0.07	.020
Economic Potential:Leader Investment:Strong Leader	−0.36	−0.56, −0.16	< .001
(Intercept)	−0.85	−1.30, −0.41	< .001
Obs.		6766	
N		273	
Group Num.		56	
DIC		12433.28	

Table S40. The coefficients in Figure 1 of the main text are derived from this MCMC GLMM regression on the z-score of contribution. The coefficients of interest can be calculated by changing the reference groups, changing the meaning of the "main effects" in the model. For example, the the coefficient of bribery game in this table is the difference between the BG treatment and the IPGG when Economic Potential and Strong Leader are zero.

Frequentist

	Coefficient	95% CI	p-value
Economic Potential	0.80	0.36, 1.23	.001
Bribery Game	−0.21	−0.31, −0.11	< .001
Full Transparency	−0.20	−0.29, −0.10	< .001
Leader Investment	−0.46	−0.56, −0.36	< .001
Partial Transparency	−0.31	−0.40, −0.22	< .001
Strong Leader	0.19	−0.25, 0.63	.432
Period	0.01	0.00, 0.01	.002
Version	0.38	0.08, 0.66	.013
Subjects	−0.01	−0.16, 0.15	.952
Order	0.15	0.13, 0.16	< .001
Age	0.00	−0.06, 0.06	.964
Male	0.20	0.07, 0.32	.002
Economic Potential:Bribery Game	−0.18	−0.33, −0.04	.013
Economic Potential:Full Transparency	0.04	−0.10, 0.18	.562
Economic Potential:Leader Investment	0.31	0.16, 0.45	< .001
Economic Potential:PartialTransparency	0.00	−0.13, 0.13	.950
Economic Potential:Strong Leader	−0.31	−0.93, 0.31	.344
Bribery Game:Strong Leader	−0.31	−0.46, −0.17	< .001
Full Transparency:Strong Leader	0.15	0.00, 0.29	.049
Leader Investment:Strong Leader	0.29	0.14, 0.44	< .001
Partial Transparency:Strong Leader	−0.22	−0.36, −0.08	.002
Economic Potential:Bribery Game:Strong Leader	0.14	−0.06, 0.33	.179
Economic Potential:Full Transparency:Strong Leader	−0.25	−0.45, −0.05	.016
Economic Potential:Leader Investment:Strong Leader	−0.35	−0.55, −0.15	.001
Economic Potential:Partial Transparency:Strong Leader	0.09	−0.11, 0.28	.367
(Intercept)	−0.81	−1.66, 0.04	.080
Obs.		6766	
N		273	
Group Num.		56	
R^2		0.68	

Table S41. Multilevel model regressing z-score of contribution, with random effects for participants within groups. . The variance explained by both fixed and random factors[34,35] is $R^2 = 0.68$.

Leader Investment

	Weak Leaders		Strong Leaders	
Poor Economic Potential	Control	BG	Control	BG
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***
BG + Leader Investment	-0.46***	-0.25***	-0.17**	0.36***
Rich Economic Potential	Control	BG	Control	BG
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***
BG + Leader Investment	-0.15**	0.24***	-0.21***	0.36***

Figure S19. Figure 2 from Main Text with Leader Investment included. 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 and Leader Investment further decrease contributions.

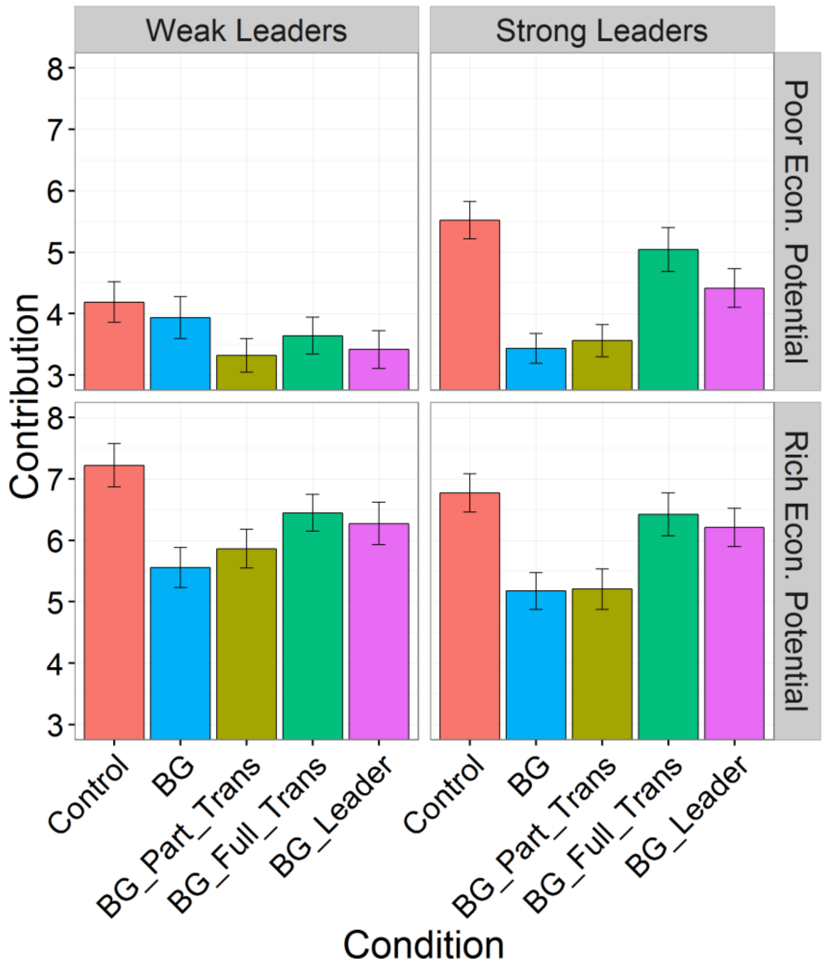


Figure S20. Figure 3 from Main Text with Leader Investment Included.

Preferences for characteristics of the game world

Questions

One more question - after having played several different versions of this game, if you were to play one more game where you chose the rules, what would you do?

In my version of the game...

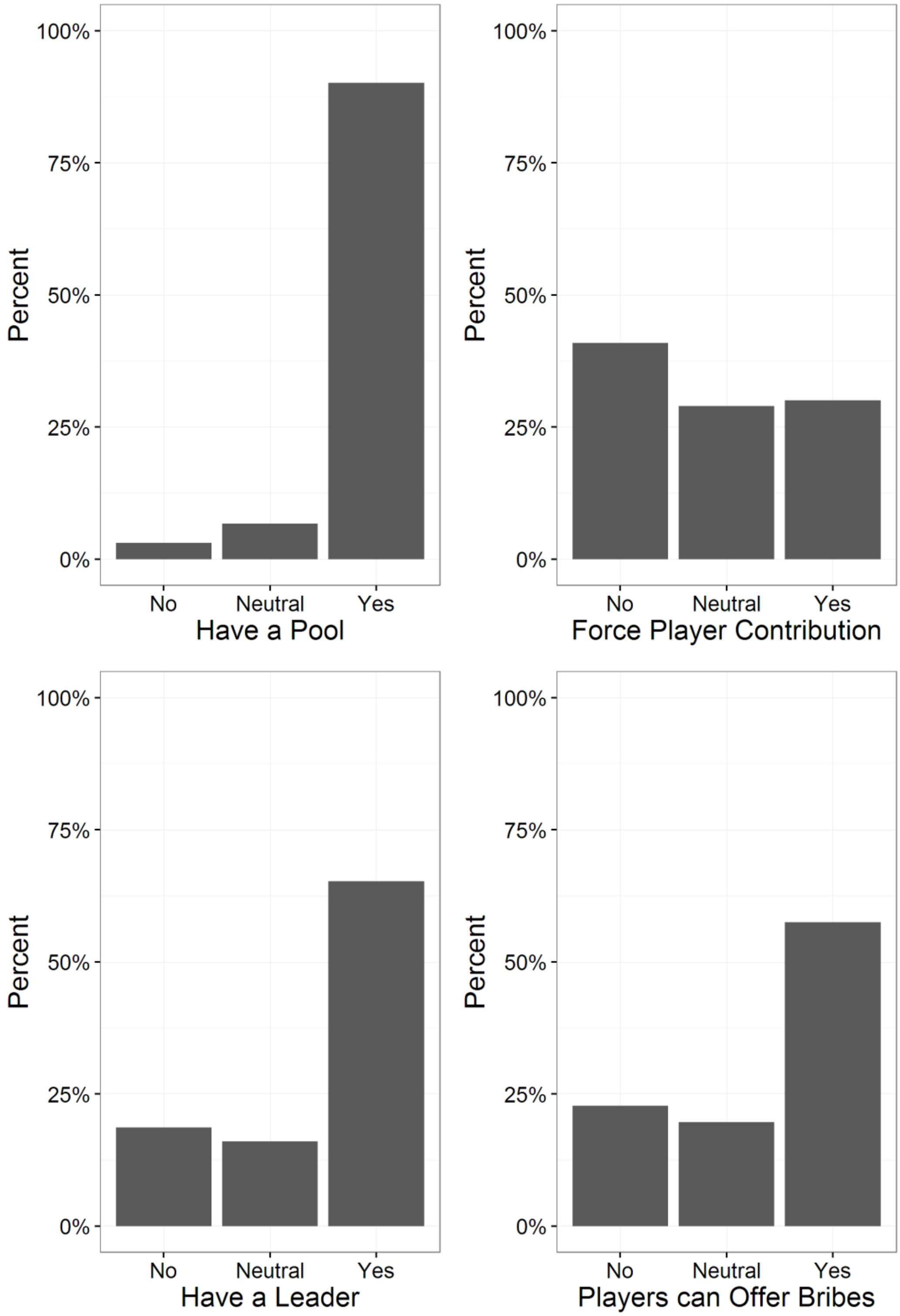
There would be a Pool				There would be NO Pool
Players would be forced to contribute 10 points				Players would NOT be forced to contribute
There would be a Leader				There would be NO Leader
The Leader could punish players				The Leader could NOT punish players
The Leader could accept payments from players				The Leader could NOT accept payments from players
The Pool Multiplier would be HIGHER				The Pool Multiplier would be LOWER
The Take Away Multiplier would be HIGHER				The Take Away Multiplier would be LOWER
The Leader would be forced to contribute 10 points				The Leader would NOT be forced to contribute
The Leader's contribution would be visible to all players				The Leader's contribution would NOT be visible to all players
Players could contribute to the Leader				Players could NOT contribute to the Leader
All Player's actions would be visible to everyone				All Player's actions would NOT be visible to everyone

We gave participants a survey at the end of the experiment to see what kind of world they would prefer were they allowed to change the parameters. We assume that this is also the kind of world they would migrate to given the opportunity. Looking only at majorities where greater than 50% agreed on something, most people want to live in a world with:

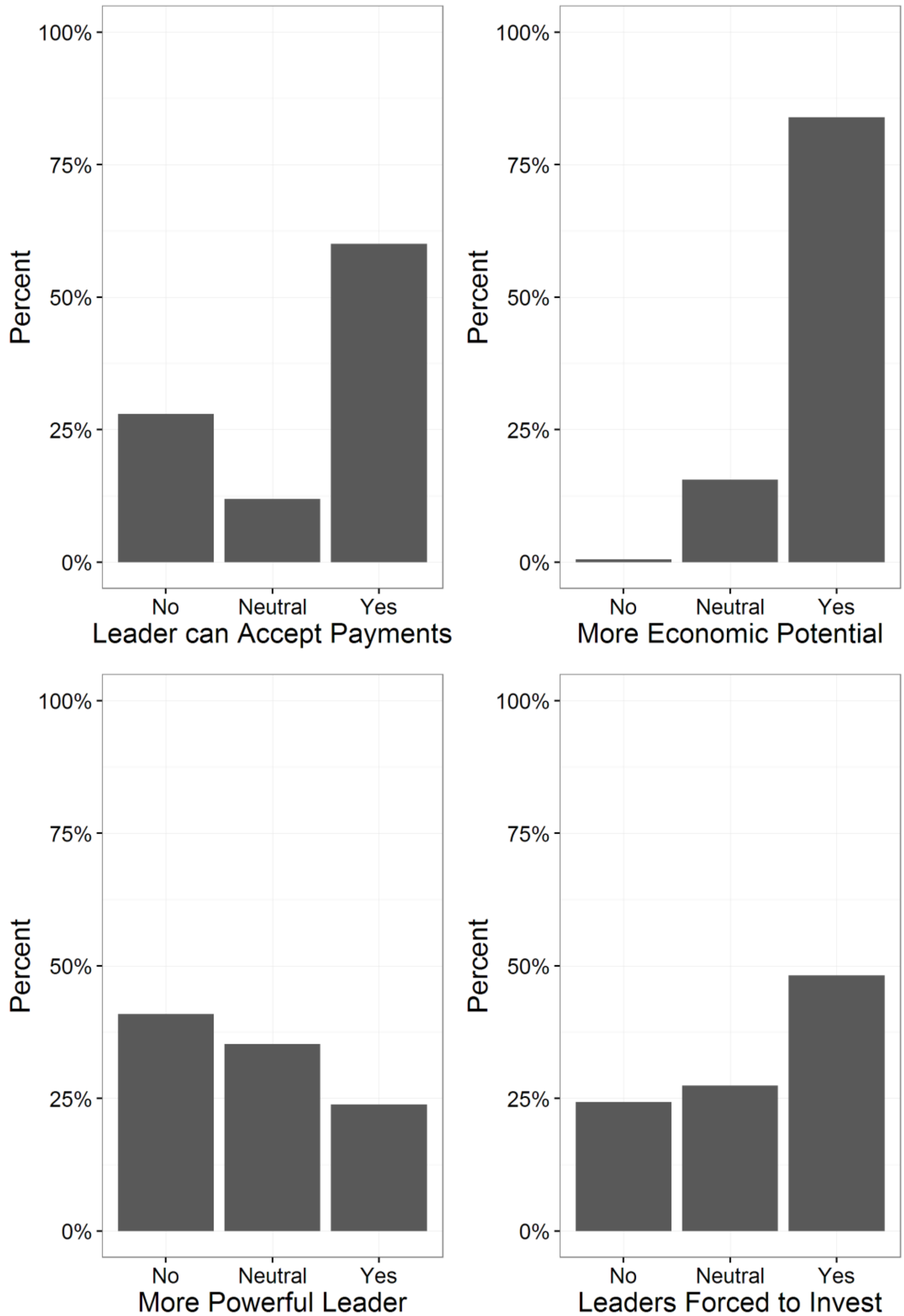
A pool with institutional punishment, but where players can offer bribes and leaders can accept these bribes. Economic potential would be rich (unsurprisingly) and there would be transparency (players expressed strong preference for both transparency types).

There is some disagreement, but a small plurality of people would prefer to choose to contribute rather than be forced to contribute, and would prefer the leader to be less powerful and forced to invest in the public good.

Graphs



100%
75%
50%
25%
0%
Percent
No
Neutral
Yes
Have a Pool
Force Player Contribution
Have a Leader
Players can Offer Bribes



Percent
100%
75%
50%
25%
0%
No
Neutral
Yes
Leader can Accept Payments
More Economic Potential
More Powerful Leader
Leaders Forced to Invest

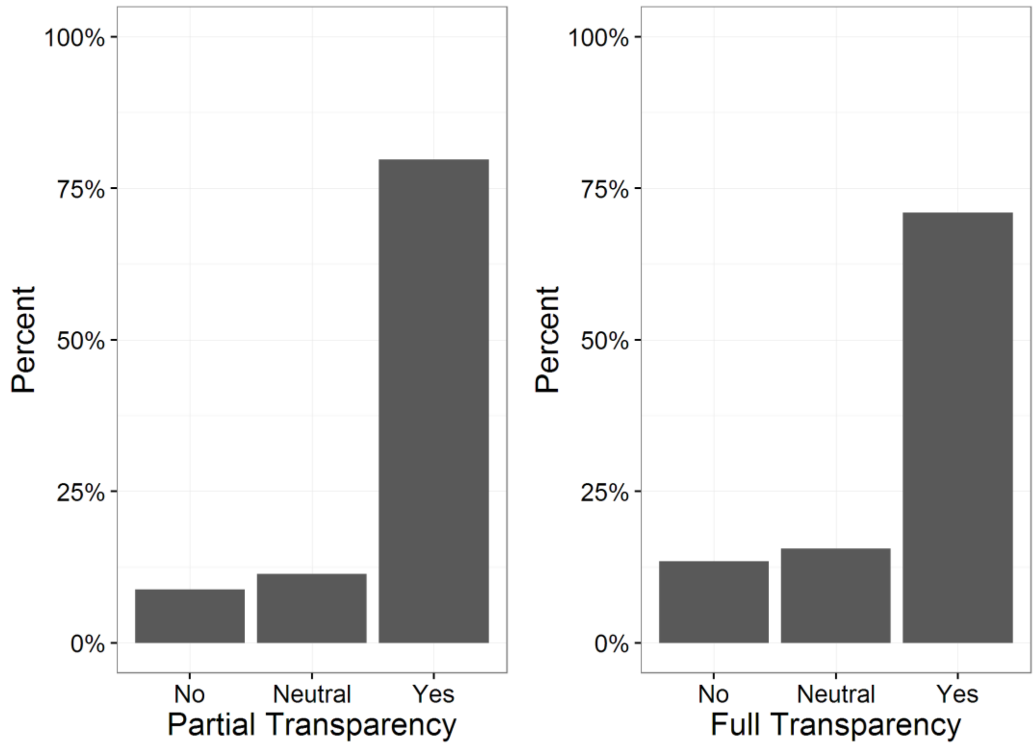


Figure S21. Distribution of answers for each end of survey question regarding preferences for the characteristics of the game.

Supplementary References

18. 1 Foltz, J. D. & Opoku-Agyemang, K. A. Do higher salaries lower petty corruption? A policy experiment on West Africa's highways. *Working Paper* (2015).
19. 2 Cheng, J. T., Tracy, J. L. & Henrich, J. Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior* **31**, 334-347 (2010).
20. 3 Altemeyer, B. *The Authoritarians*. (Lulu Press, 2006).
21. 4 Aron, A., Aron, E. N. & Smollan, D. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of personality and social psychology* **63**, 596 (1992).
22. 5 Ryder, A. G., Alden, L. E. & Paulhus, D. L. Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of personality and social psychology* **79**, 49 (2000).
23. 6 Transparency International. (2014).
24. 7 Henrich, J. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization* **53**, 3-35 (2004).
25. 8 Boyd, R. & Richerson, P. J. *Culture and the evolutionary process*. (University of Chicago Press, 1985).
26. 9 Cavalli-Sforza, L. L. & Feldman, M. W. *Cultural transmission and evolution: a quantitative approach*. (Princeton University Press, 1981).
27. 10 Doebeli, M., Hauert, C. & Killingback, T. The evolutionary origin of cooperators and defectors. *Science* **306**, 859-862 (2004).
28. 11 Page, K. M. & Nowak, M. A. A generalized adaptive dynamics framework can describe the evolutionary Ultimatum Game. *Journal of Theoretical Biology* **209**, 173-179 (2001).

29. 12 Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W. & Uller, T. Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? *science* **334**, 1512-1516 (2011).
30. 13 Rendell, L. *et al.* Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in cognitive sciences* **15**, 68-76 (2011).
31. 14 Horner, V. & Whiten, A. Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal cognition* **8**, 164-181 (2005).
32. 15 McGuigan, N., Whiten, A., Flynn, E. & Horner, V. Imitation of causally opaque versus causally transparent tool use by 3-and 5-year-old children. *Cognitive Development* **22**, 353-364 (2007).
33. 16 Henrich, J. & McElreath, R. The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews* **12**, 123-135 (2003).
34. 17 Bell, A. V. Linking observed learning patterns to the evolution of cultural complexity. *Current Anthropology* **56**, 277-281 (2015).
35. 18 Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E. & Kendal, R. L. Human cumulative culture: a comparative perspective. *Biological Reviews* **89**, 284-301 (2014).
36. 19 Enquist, M., Strimling, P., Eriksson, K., Laland, K. & Sjostrand, J. One cultural parent makes no culture. *Animal Behaviour* **79**, 1353-1362 (2010).
37. 20 Besley, T. & Persson, T. Democratic Values and Institutions. *Working Paper* (2016).
38. 21 Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785-791 (2003).
39. 22 Henrich, J. *et al.* In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 73-78 (2001).
40. 23 Tabellini, G. The Scope of Cooperation: Values and Incentives. *The Quarterly Journal of Economics*, 905-950 (2008).
41. 24 Francois, P. & Zabojnik, J. Trust, social capital, and economic development. *Journal of the European Economic Association* **3**, 51-94 (2005).
42. 25 Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61-83 (2010).
43. 26 Mesoudi, A. How cultural evolutionary theory can inform social psychology and vice versa. *Psychological Review* **116**, 929-952 (2009).
44. 27 Richerson, P. *et al.* Cultural Group Selection Plays an Essential Role in Explaining Human Cooperation: A Sketch of the Evidence. *Behavioral and Brain Sciences* **39** (2016).
45. 28 Henrich, J. & Gil-White, F. J. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior* **22**, 165-196 (2001).
46. 29 Chudek, M. & Henrich, J. Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences* **15**, 218-226 (2011).
47. 30 Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* **33**, 1-22 (2010).
48. 31 Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472 (1992).
49. 32 Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: Convergence diagnosis and output analysis for MCMC. *R news* **6**, 7-11 (2006).
50. 33 Hadfield, J. (2015).

51. 34 Johnson, P. C. D. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* **5**, 944-946, doi:10.1111/2041-210X.12225 (2014).
52. 35 Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133-142, doi:10.1111/j.2041-210x.2012.00261.x (2013).