Online Appendix for "Does Trust in Government Increase Support for Redistribution? Results from Randomized Survey Experiments"

Kyle Peyton*

Version: January 2020

Contents

*Postdoctoral Fellow in Law and Social Science, The Justice Collaboratory, Yale Law School. Email: kyle.peyton@yale.edu.

List of Figures

List of Tables

S1 Overview

Table S1 summarizes the five survey experiments, conducted between June 2014 and March 2017, described in the manuscript. Excepting Experiment 2, participants were recruited from Amazon Mechanical Turk (MTurk) and assigned to one of three treatment arms – *Corrupt*, *Control*, or *Honest* – with equal probability according to simple random assignment. Experiment 2, conducted on a nationally representative online panel, deviated from the simple design and used block random assignment by party identification to ensure balance in partisanship across treatment conditions. All estimators use inverse probability weighting (IPW) to account for the different assignment probabilities in Experiment 2.

Figure S1 illustrates the basic design of each experiment. All experiments were conducted using Qualtrics survey software. Pre-treatment questions included standard demographic measures and party identification. The *Honest* and *Corrupt* treatments were putative Op-Eds published in *The New York Times*, inspired by a real *New York Times* article (see Wines, 2014) about political corruption. The Op-Eds in Experiments 1-2 were identical. In *Honest*, participants read an Op-Ed titled "'It Only Seems that Political Corruption is Rampant" that emphasized the integrity of government officials and low levels of corruption. In *Corrupt*, participants read an Op-Ed, titled "Political Corruption is Rampant", that used contrasting language about the lack of integrity among government officials and the prevalence of political corruption. In *Control*, participants read a piece about celebrity chef Anthony Bourdain.

Experiment 3 also provided data visualizations to support the Op-Ed writer's argument, but the content otherwise matched Experiments 1-2. *Control* in Experiment 3 was an Op-Ed about recycling, also supplemented with a data visualization. *Control* in Experiments 4-5 were identical to Experiments 1-2, and the *Honest* and *Corrupt* treatment arms provided information about the absence or presence of corruption in the National Football League (NFL) instead of politics. See Section A for the full text that appeared in each treatment arm across Experiments 1-5 and additional details.

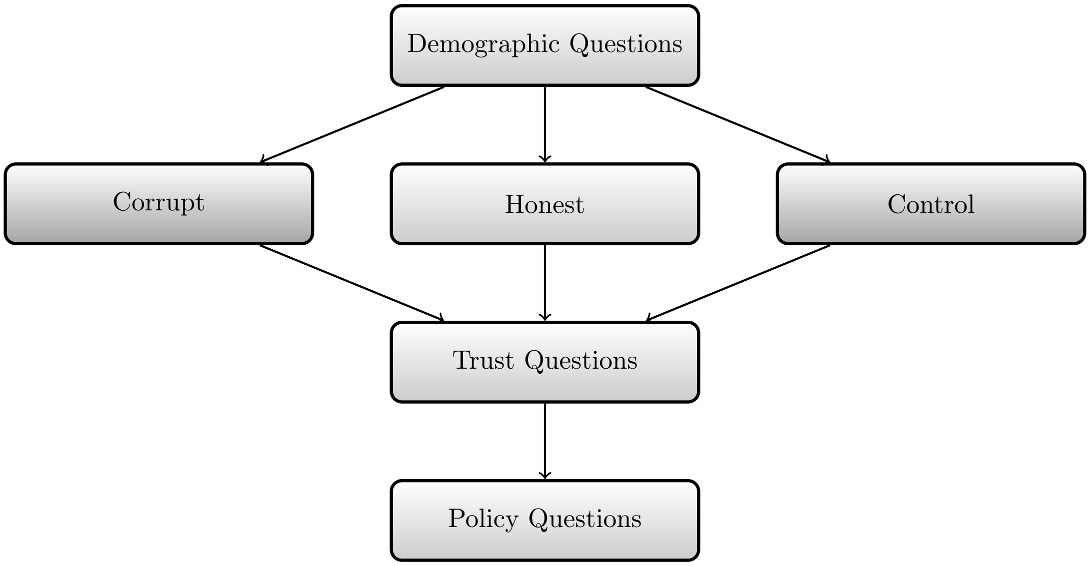
All of the Op-Eds in Experiments 1-5 were loosely based on past events and drew on factual information about real cases of misconduct in American politics and the NFL; however, they were not real Op-Eds and therefore used mild deception. This was approved by the Human Ethics committee in Experiments 1, 2 and 4, conditional on a de-brief at the end of the survey with an option for participants to remove their data. Figure S10 shows this prompt. All subjects that elected to have their data removed are excluded from analyses per IRB requirements. There is no evidence that these decisions were a function of treatment assignment in any of experiment (see Section S2-S3). The author had moved to a different university when Experiments 3 and 5 were designed and implemented. The Institutional Review Board (IRB) at this university also required a debrief, but did not mandate that subjects also be given the opportunity to remove their data afterward.

Section S2 provides additional details about the sample characteristics for Experiments 1-3, along with tests of key design assumptions. Section S3 provides this information for Experiments 4-5. Section S5 provides a variety of additional analyses, robustness checks, and explanatory notes, including details about subject recruitment, survey-taker attentiveness, and an overview of recent research on the nature of "demand effects" in online survey experiments.

TABLE S1: Summary of Experiments 1-5

	Op-Ed Content	Date	Platform	Sample Size
Experiment 1:	Political Corruption	June 2014	MTurk	643
Experiment 2:	Political Corruption	September 2014	Qualtrics Panels	1324
Experiment 3:	Political Corruption	March 2017	MTurk	1870
Experiment 4:	Non-Political Corruption	December 2014	MTurk	584
Experiment 5:	Non-Political Corruption	July 2015	MTurk	585

FIGURE S1: Experiment flow diagram



S2 Political Corruption Experiments

S2.1 Experiment 1 (June 2014)

692 subjects were recruited via MTurk and 20 respondents were excluded for having the same IP Address. Of the remaining 672 subjects, 29 asked to have their data removed from the experiment after learning about deception in the debrief. These responses were removed in accordance with IRB requirements. An F-test from a linear regression of removal on treatment assignment confirmed that assignment was not predictive of the removal request (P-value $= 0.78$). Table S2 shows the allocation of the remaining 643 subjects across treatment arms. Table S3 shows sample characteristics. Randomization inference is used to asses covariate balance across treatment arms (see Gerber and Green, 2012, Chapter 4 for a textbook treatment) using the `ri2` package in `R` (Coppock, 2018). Figure S2 plots a histogram of the observed F-statistic, and the null distribution of F-statistics, from a regression of treatment assignment on covariates. Approximately 39% of the simulated F-statistics were larger than the observed F-statistic (P-value of 0.39). Thus, the null hypothesis that no covariates have any effect on treatment assignment, as implied by the experimental design, is not rejected.

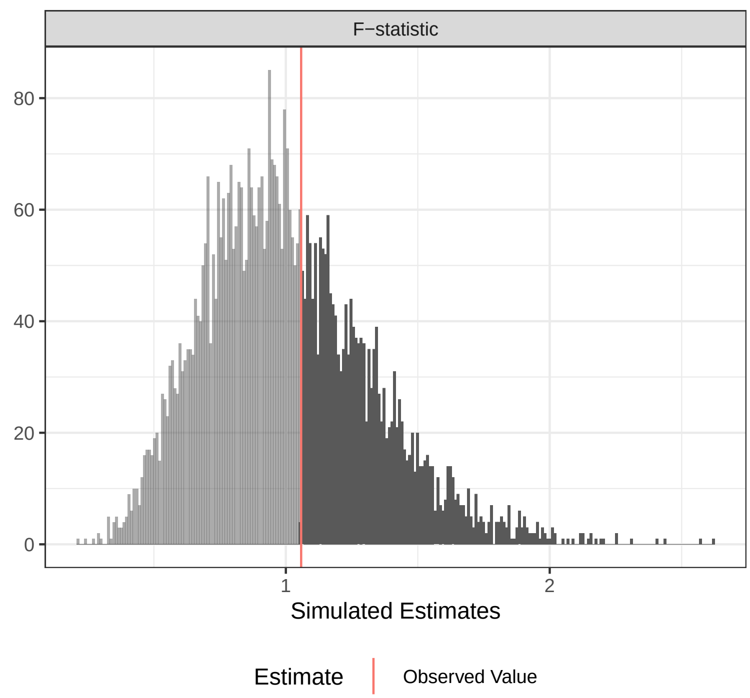
TABLE S2: Treatment Assignments in Experiment 1

	Treatment Condition			
	Control	Corrupt	Honest	Totals
N	215	217	211	643

Table S3: Sample characteristics in Experiment 1

	Mean	St. Dev.	Median	Min	Max
Female	0.46	0.50	0.00	0	1
Age	35.37	11.82	32.00	19	80
College degree	0.49	0.50	0.00	0	1
Employed	0.78	0.42	1.00	0	1
Asian	0.05	0.22	0.00	0	1
Black	0.08	0.28	0.00	0	1
Hispanic	0.04	0.20	0.00	0	1
White	0.79	0.40	1.00	0	1
Democrat	0.44	0.50	0.00	0	1
Republican	0.18	0.39	0.00	0	1

Figure S2: Randomization Inference for Covariate Balance in Experiment 1



Notes: The vertical red line denotes the observed F-statistic (1.06). Shaded regions denote simulated estimates more extreme than the one observed. The randomization inference P-value is 0.39. The test is based on 5,000 simulations under the null hypothesis that no covariates have any effect on treatment assignment, as implied by random assignment.

S2.2 Experiment 2 (Sept. 2014)

1474 subjects were recruited via Qualtrics Panels and 1452 remained after excluding duplicate IP Addresses. Of these, 128 asked to have their data removed from the experiment after learning about deception. An F-test from a linear regression of removal on treatment assignment confirmed that assignment was not predictive of the removal request (P-value = 0.38). The final sample size was 1324. This was a US General population sample and used the following quotas:

1. Age: Atleast 86% of sample less than 65 years old
2. Sex: 50/50 balance
3. Race: 63% of the population should be white, 13% should be black and 17% should be Hispanic.
4. Education: 85% High School or higher. At least 28% bachelors degree or higher.
5. Party identification: 42% should be Independent; 25% Republican; 31% Democrat

MTurk workers tend to skew white, educated, and liberal (see Berinsky, Huber and Lenz, 2012). The party identification quotas were chosen in light of contemporaneous Gallup polls showing the increasing proportion of self-identified Independents in the United States[1], and the race quotas were chosen based on 2013 census estimates.[2] All quotas were approximately met, so this sample is a reasonable approximation to a nationally representative sample of Americas on these observables.

Table S4 shows the allocation of subjects across conditions. Table S5 shows sample characteristics. Randomization inference is again used to asses covariate imbalance across

[1]see: http://www.gallup.com/poll/180440/new-record-political-independents.aspx
[2]see http://quickfacts.census.gov/qfd/states/00000.html

treatment arms. Figure S3 plots a histogram of the observed F-statistic, and the null distribution of F-statistics, from a regression of treatment assignment on covariates. Approximately 73% of the simulated F-statistics were larger than the observed F-statistic (P-value of 0.73). Thus, the null hypothesis that no covariates have any effect on treatment assignment cannot be rejected, as implied by the experimental design.

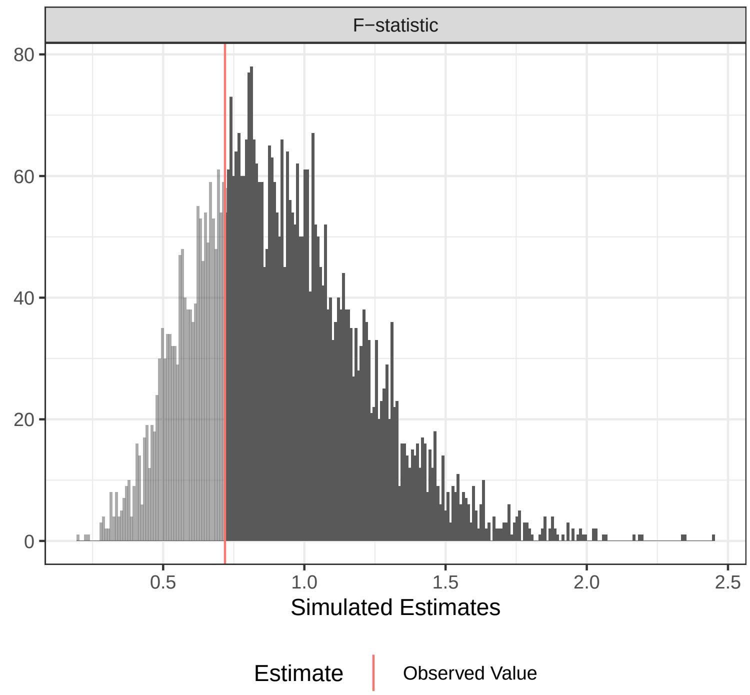
Table S4: Treatment Assignments in Experiment 2

	Treatment Condition			
	Control	Corrupt	Honest	Totals
N	430	443	451	1324

Table S5: Sample characteristics in Experiment 2

	Mean	St. Dev.	Median	Min	Max
Female	0.49	0.50	0.00	0	1
Age	43.38	16.32	41.00	18	88
College degree	0.41	0.49	0.00	0	1
Employed	0.60	0.49	1.00	0	1
Asian	0.05	0.22	0.00	0	1
Black	0.13	0.33	0.00	0	1
Hispanic	0.17	0.37	0.00	0	1
White	0.64	0.48	1.00	0	1
Democrat	0.31	0.46	0.00	0	1
Republican	0.27	0.45	0.00	0	1

Figure S3: Randomization Inference for Covariate Balance in Experiment 2



Notes: The vertical red line denotes the observed F-statistic (0.72). Shaded regions denote simulated estimates more extreme than the one observed. The randomization inference P-value is 0.73. The test is based on 5,000 simulations under the null hypothesis that no covariates have any effect on treatment assignment, as implied by random assignment.

S2.3 Experiment 3 (March 2017)

1976 subjects were recruited via MTurk and 106 were excluded for attempting to access the survey more than once from the same IP Address for a final sample of 1870 subjects. Table S6 shows the allocation of these subjects across treatment arms. Table S7 shows samples characteristics. Randomization inference is again used to asses covariate imbalance across treatment arms. Figure S4 plots a histogram of the observed F-statistic, and the null distribution of F-statistics, from a regression of treatment assignment on covariates. Approximately 55% of the simulated F-statistics were larger than the observed F-statistic (P-value of 0.55). Thus, the null hypothesis that no covariates have any effect on treatment assignment is not rejected, as implied by the experimental design.

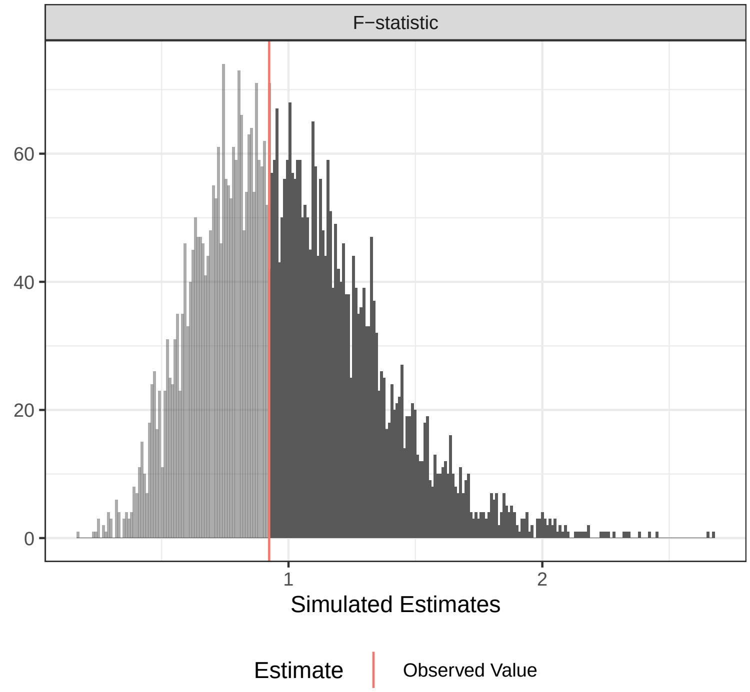
Table S6: Treatment Assignments in Experiment 3

	Treatment Condition			
	Control	Corrupt	Honest	Totals
N	662	583	625	1870

Table S7: Sample characteristics in Experiment 3

	Mean	St. Dev.	Median	Min	Max
Female	0.46	0.50	0.00	0	1
Age	36.33	11.00	34.00	18	72
College degree	0.46	0.50	0.00	0	1
Employed	0.74	0.44	1.00	0	1
Asian	0.07	0.26	0.00	0	1
Black	0.08	0.27	0.00	0	1
Hispanic	0.05	0.21	0.00	0	1
White	0.75	0.43	1.00	0	1
Democrat	0.44	0.50	0.00	0	1
Republican	0.22	0.41	0.00	0	1

Figure S4: Randomization Inference for Covariate Balance in Experiment 3



Notes: The vertical red line denotes the observed F-statistic (0.92). Shaded regions denote simulated estimates more extreme than the one observed. The randomization inference P-value is 0.55. The test is based on 5,000 simulations under the null hypothesis that no covariates have any effect on treatment assignment, as implied by random assignment.

S3 Placebo Experiments

The exclusion restriction assumption, as explained in the manuscript, holds that any treatment effect on support for redistribution occurs through increased political trust, which could be violated if treatment impacted support for redistribution via another pathway. Two survey experiments randomly assigned participants to receive information about corruption in the National Football League (NFL). These "placebo experiments" are used to investigate whether the increased political trust in Experiments 1-3 is simply a function of content valence, or priming respondents to think about corruption in a salient non-government institution[3] If treatment effects on political trust in Experiments 1-3 are simply a function of content valence, then an effect should also be detectable when political content is removed. If support for redistribution – *but not political trust* – is affected by treatments about non-political corruption, this provides evidence of an ER violation and suggests an alternative mechanism could bias results from Experiments 1-3.

S3.1 Experiment 4 (December 2014)

624 subjects were recruited using MTurk and 612 remained after excluding duplicate IP Addresses. Of those remaining, 28 asked to have their data removed from the experiment after learning about deception in the debrief. An F-test from a linear regression of removal on treatment assignment confirmed that assignment was not predictive of the removal request (P-value = 0.35). The final sample size was 584. Table S8 shows the allocation of subjects across conditions. Table S9 shows demographic characteristics. Randomization inference is used to asses covariate imbalance across treatment arms. Figure S5 plots a histogram of the

[3] One way to asses the NFL's popularity is to look at viewership statistics. Publicly available estimates from Harris Interactive suggest football is, by a wide margin, the most popular professional sport in the United States (Interactive, 2015). According to data published by RBC Capital Markets (Ciolli, 2017), the estimated NFL season TV audience was approximately 15 million persons per game in 2017, down by about 12% from the roughly 17 million persons per game when the NFL experiments were conducted in 2014 and 2015.

observed F-statistic, and the null distribution of F-statistics, from a regression of treatment assignment on covariates. Approximately 83% of the simulated F-statistics were larger than the observed F-statistic (P-value of 0.83). Thus, the null hypothesis that no covariates have any effect on treatment assignment in not rejected, as implied by the experimental design.

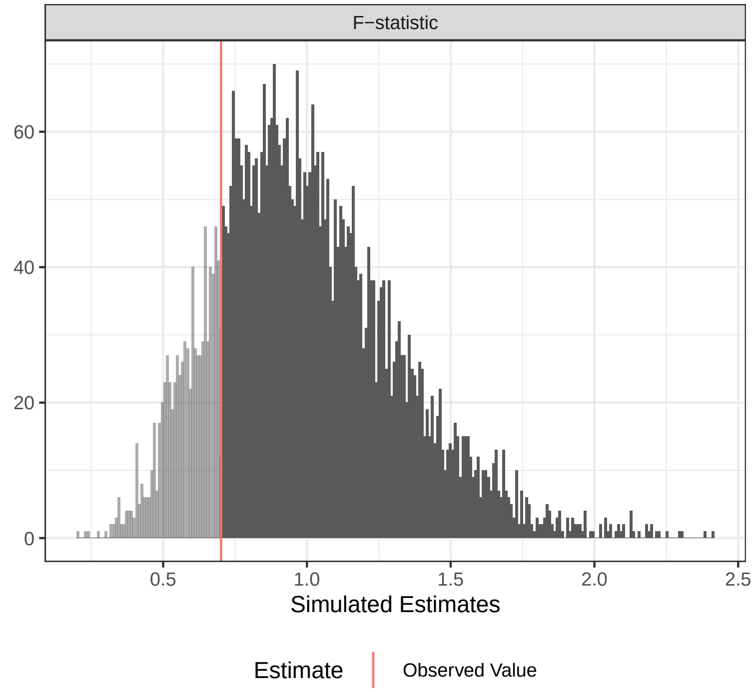
Table S8: Treatment Assignments in Experiment 4

	Treatment Condition			
	Control	Corrupt	Honest	Totals
N	194	195	195	584

Table S9: Sample characteristics in Experiment 4

	Mean	St. Dev.	Median	Min	Max
Female	0.48	0.50	0.00	0	1
Age	33.46	11.16	30.00	18	84
College degree	0.47	0.50	0.00	0	1
Employed	0.79	0.41	1.00	0	1
Asian	0.09	0.28	0.00	0	1
Black	0.07	0.26	0.00	0	1
Hispanic	0.07	0.25	0.00	0	1
White	0.74	0.44	1.00	0	1
Democrat	0.46	0.50	0.00	0	1
Republican	0.15	0.35	0.00	0	1

Figure S5: Randomization Inference for Covariate Balance in Experiment 4



Notes: The vertical red line denotes the observed F-statistic (0.70). Shaded regions denote simulated estimates more extreme than the one observed. The randomization inference P-value is 0.83. The test is based on 5,000 simulations under the null hypothesis that no covariates have any effect on treatment assignment, as implied by random assignment.

S3.2 Experiment 5 (July 2015)

612 subjects were recruited using MTurk and 27 were excluded for attempting to access the survey more than once from the same IP Address for a final sample of 585. Table S10 shows the allocation of subjects across conditions. Table S11 shows other demographic characteristics. Randomization inference is again used to asses covariate imbalance across treatment arms. Figure S6 plots a histogram of the observed F-statistic, and the null distribution of F-statistics, from a regression of treatment assignment on covariates. Approximately 44% of the simulated F-statistics were larger than the observed F-statistic (P-value of 0.44). Thus, the null hypothesis that no covariates have any effect on treatment assignment is not rejected,

as implied by the experimental design.

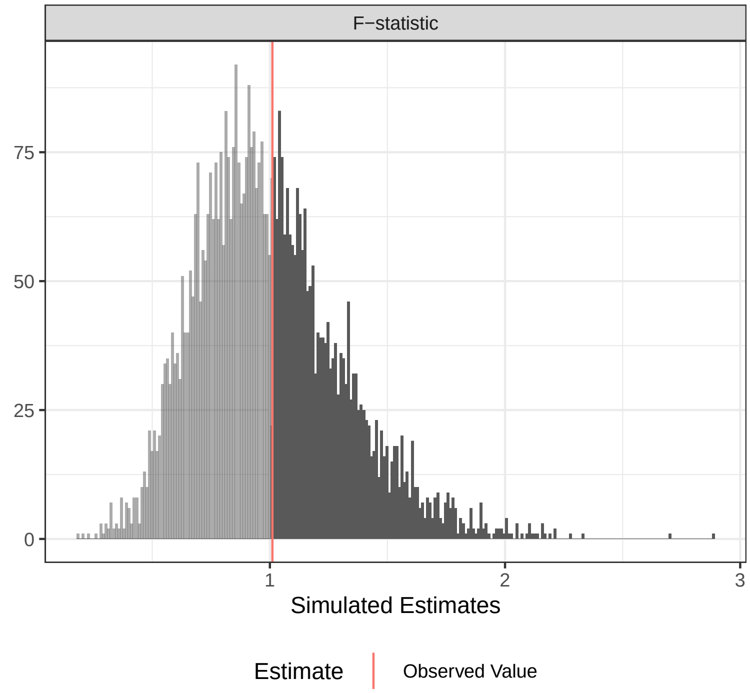
Table S10: Treatment Assignments in Experiment 5

	Treatment Condition			
	Control	Corrupt	Honest	Totals
N	211	176	198	585

Table S11: Sample characteristics in Experiment 5

	Mean	St. Dev.	Median	Min	Max
Female	0.55	0.50	1.00	0	1
Age	34.74	11.64	32.00	19	78
College degree	0.52	0.50	1.00	0	1
Employed	0.77	0.42	1.00	0	1
Asian	0.06	0.24	0.00	0	1
Black	0.08	0.27	0.00	0	1
Hispanic	0.05	0.22	0.00	0	1
White	0.77	0.42	1.00	0	1
Democrat	0.41	0.49	0.00	0	1
Republican	0.18	0.38	0.00	0	1

Figure S6: Randomization Inference for Covariate Balance in Experiment 5



Notes: The vertical red line denotes the observed F-statistic (1.01). Shaded regions denote simulated estimates more extreme than the one observed. The randomization inference P-value is 0.44. The test is based on 5,000 simulations under the null hypothesis that no covariates have any effect on treatment assignment, as implied by random assignment.

S4 Outcome measurement

This section describes the survey instruments, question wording and coding of responses, and the sample means (and standard errors) across treatment arms for the measures of political trust and support for redistribution used in Experiments 1-5. Reliability estimates for the composite measures of trust in government and support for redistribution are also presented for respondents assigned *Control* in each experiment.

S4.1 Trust in Government

All experiments use two measures of political trust. The primary measure is a Likert Scale from Faulkner, Martin and Peyton (2015), and the secondary measure is the single-item trust in government measure that has appeared in the ANES survey since 1958, as well as various public opinion polls conducted by Gallup and Pew. Given concerns about attenuation bias and measurement error in the ANES measure (more below), and high internal reliability of the Likert Scale, it is the preferred measures of political trust. The main results reported in the manuscript do not depend on which measure is used. As one would expect (see Ansolabehere, Rodden and Snyder, 2008), estimates based on the ANES item are less precise than the additive scale. As one would expect, political trust is low in *Control* across all experiments, regardless of which measure is preferred. Trust in government has been at historic lows in the United States over the past decade, and the low levels of political trust observed in the survey experiments reported here are consistent with survey data from a variety of other sources (i.e. Pew, Gallup, ANES).

S4.1.1 Likert Scale

Respondents were asked to indicate their agreement with each of the following statements measured on a 6 point scale from “Strongly Disagree” to “Strongly Agree”:

1. We generally cannot trust politicians.
2. People in government are too often interested in looking after themselves.
3. Government is run by a few big interests who look after their own interests.
4. A lot of politicians are corrupt.

An additive scale (with range 1-24) was constructed by summing the responses to these 4 items. Figure S7 shows the response distribution in *Control* across all five experiments. The Omega statistics for internal consistency (Dunn, Baguley and Brunsden, 2013), calculated using the MBESS packaged in `R` (Kelley and Lai, 2012), are reported in Table S12 (with 95% confidence intervals).

Figure S7: Response distribution in Control Group for Likert Scale

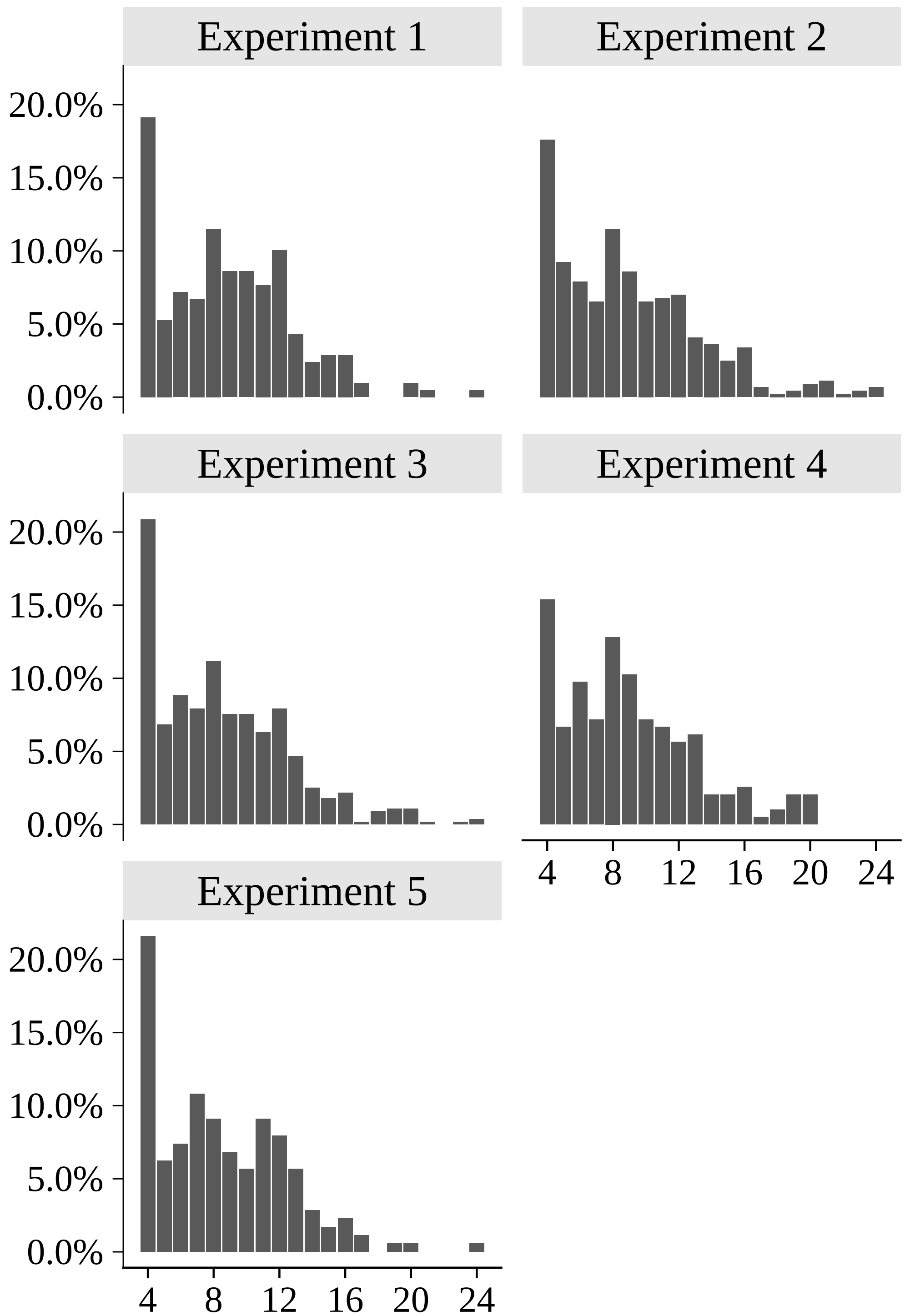


Table S12: Internal reliability of Likert Scale

Est.	SE	LB	UB	N (Control)	Experiment
0.90	0.02	0.87	0.93	217	Experiment 1
0.91	0.01	0.89	0.93	443	Experiment 2
0.91	0.01	0.89	0.92	583	Experiment 3
0.92	0.01	0.90	0.94	195	Experiment 4
0.92	0.01	0.90	0.95	176	Experiment 5

S4.1.2 ANES Item

Since 1958, the ANES survey[4] has measured trust in government with the following prompt:

> "People have different ideas about the government in Washington. These ideas don't refer to democrats or republicans in particular, but just to government in general. We want to see how you feel about these ideas."

> "How much of the time do you think you can trust the government in Washington to do what is right– just about always, most of the time or only some of the time?"

Responses are recorded on a three point scale: just about always (1), most of the time (2), only some of the time (3). Survey respondents can also volunteer a response of "never" (4). This three item scale is the subject of frequent criticism by users of the ANES, and changes have been proposed in ANES pilot reports (e.g. Gershtenson and Plane, 2007). In the 2012 version of the ANES Survey, respondents were randomly assigned to either the standard version, or an alternative version with a different response scale.[5] Respondents assigned

[4]See http://www.electionstudies.org/

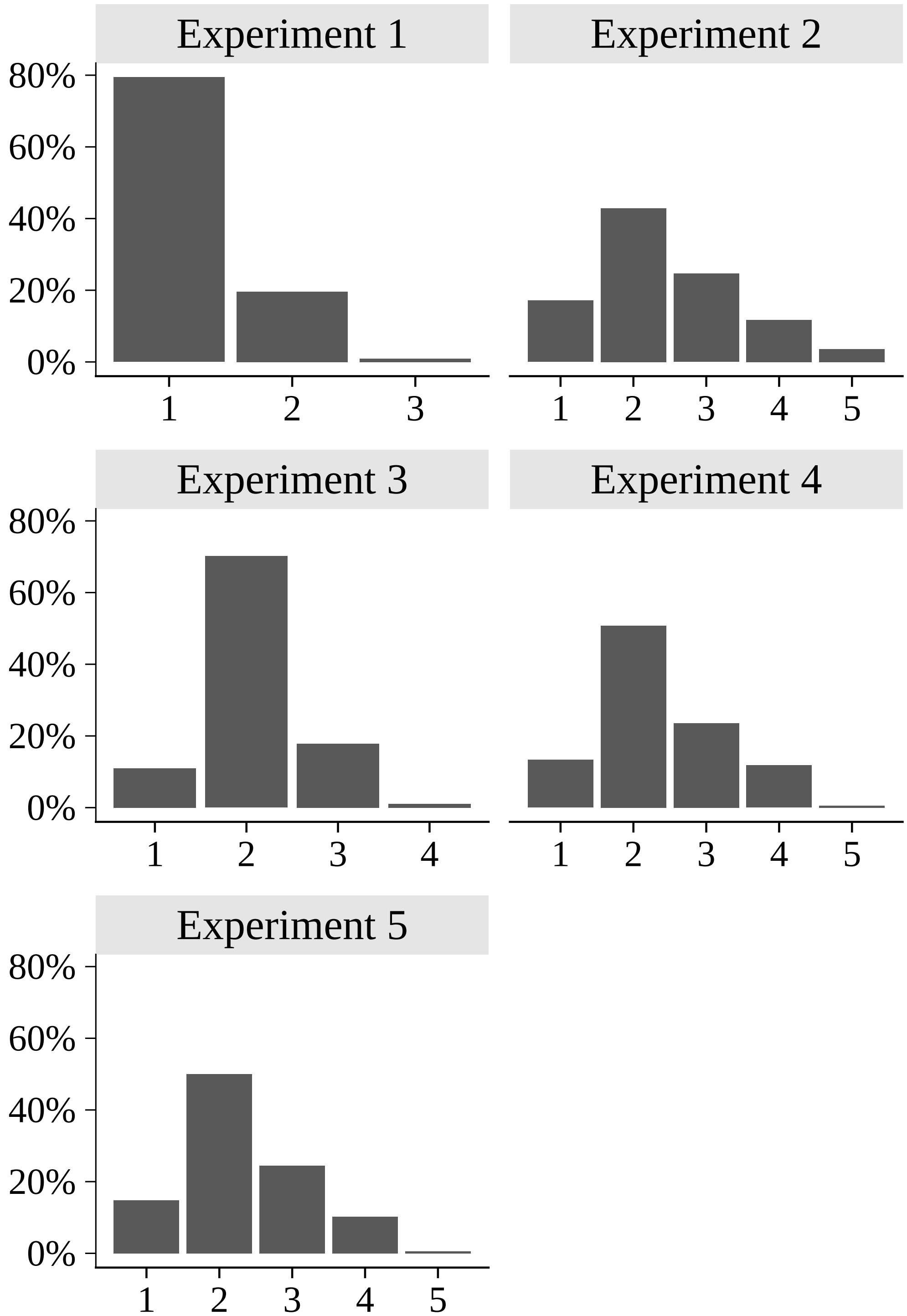
[5]See http://electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012_userguidecodebook.pdf

to this alternative version were asked the same question but given the following response options: Always (1), Most of the time (2), About half the time (3), Some of the time (4), Never (5). The standard three-item measure of political trust was used in Experiment 1. Experiments 2, 4, and 5 used this revised 5-item measure. Experiment 3 used the four-item measure from Kuziemko et al. (2015): Never (1), Some of the time (2), Most of the time (3), Just about always (4).

All question responses are rescaled so that higher values correspond to more trust in government. As expected, the number of response options affects the distribution of responses. This variation is illustrated graphically in Figure S8. In Experiment 1, 79% of respondents in *Control* said they trusted government "only some of the time", and only 2 respondents said they trusted government "always". This is consistent with recent estimates from nationally representative surveys such as the ANES Time Series Survey[6], as well as similar public opinion polls conducted by Pew and Gallup that use the same question format (Doherty et al., 2015). When additional response options were available, however, the proportion of respondents selecting the most extreme level of "distrust" decreased substantially. In Experiment 2, only 43% of respondents in *Control* indicated they trusted the government "some of the time". Results were similar for both placebo experiments – 52% of respondents in Experiment 4 and 49% in Experiment 5. Finally, results from Experiment 3 suggest removing the midpoint "about half the time" also has a non-trivial impact on the proportion of respondents selecting "some of the time". In this experiment, approximately 70% of respondents in *Control* reported they trust the government "some of the time". These large differences in reported levels of trust in government – largely attributable to question format choices – suggest that low levels of political trust reported in many nationally representative surveys that use the 3-item format may be misleading.

[6]See http://www.electionstudies.org/nesguide/toptable/tab5a_1.htm. In 2012, 76% of respondents said they trusted government "some of the time".

Figure S8: Response distribution in Control Group for ANES Item



S4.2 Support for Redistribution

All experiments used four questions about support for redistributive social policy that have appeared in various versions of the ANES Survey since 1984, and have been used in foundational studies of political trust and support for redistribution (e.g. Hetherington, 2005). Respondents are asked "Should federal spending on X be decreased (0), kept the same (0.5), or increased (1)?"

1. Food stamps: decrease (0), remain the same (0.5) or increase (1)
2. Welfare programs: decrease (0), remain the same (0.5) or increase (1)
3. Programs that assist blacks and other minorities: decrease (0), remain the same (0.5) or increase (1)
4. Programs that assist the homeless: decrease (0), remain the same (0.5) or increase (1)

Responses to these items are highly correlated and were combined to create an additive scale to reduce measurement error (see Ansolabehere, Rodden and Snyder, 2008). The Omega statistics for internal consistency (Dunn, Baguley and Brunsden, 2013), calculated using the MBESS packaged in `R` (Kelley and Lai, 2012), are reported in Table S12 (with 95% confidence intervals).

TABLE S13: Internal reliability of Support for Redistribution Index

Est.	SE	LB	UB	N (Control)	Experiment
0.87	0.01	0.85	0.90	217	Experiment 1
0.85	0.01	0.83	0.88	443	Experiment 2
0.86	0.01	0.84	0.88	583	Experiment 3
0.88	0.02	0.84	0.91	195	Experiment 4
0.89	0.01	0.86	0.92	176	Experiment 5

S5 Supplementary analyses and explanatory notes

Glass's Δ: Unless otherwise indicated, all results presented in this section, and in the manuscript, are standardized using Glass's Δ (Glass, 1976). This is simply a linear transformation of the raw data commonly used to facilitate interpretation: treatment effects are expressed in terms of a standardized effect size (see also Gerber and Green, 2012, pp. 70-71). To illustrate, let Y_i denote some response for individual i, and $Z_i = \{1, 2, 3\}$ denote treatment assignment so that $Z_i = 1$ if assigned *Control*, $Z_i = 2$ if assigned *Honest*, and $Z_i = 3$ if assigned *Corrupt*. Suppose N_1 units are assigned *Control*, N_2 units are assigned *Honest* and N_3 units are assigned *Corrupt*. Within each experiment, responses scaled by Glass's Δ are calculated by dividing by the standard deviation of the response in *Control* so that the rescaled response, $\tilde{Y}_i$, is

$$\tilde{Y}_i = Y_i \cdot \left(\sqrt{\frac{\sum_{i \in N_1} (Y_i - \bar{Y})^2}{N_1 - 1}} \right)^{-1}$$

Where $\bar{Y}$ denotes the control group mean.

IV Estimation: I estimate the effect of political trust on support for redistribution using Two-Stage Least Squares (2SLS). Equation 1 models subject i's support for redistribution, Y_i, endogenous political trust, T_i, pre-treatment covariates, $X_{1i}, X_{2i}, \dots, X_{Ki}$, and unmeasured factors U_i.

$$Y_i = \beta_0 + \beta_1 T_i + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \cdots + \gamma_K X_{Ki} + U_i \tag{1}$$

If $\text{Cov}(T_i, U_i) = 0$ then β_1 can be consistently estimated using ordinary least squares (OLS) regression. In the general case where T_i is endogenous, as in prior studies that have relied on regression adjustment, estimates from OLS are biased and inconsistent without further untestable assumptions.

The "first-stage" estimating equation specifies a linear additive relationship between political trust, the randomly assigned instrument (Z_i is a three level factor for treatment assignment: $\mathit{Corrupt} = 0$, $\mathit{Control} = 0.5$, or $\mathit{Honest} = 1$) and pre-treatment covariates:

$$T_i = \alpha_0 + \alpha_1 Z_i + \delta_1 X_{1i} + \delta_2 X_{2i} + \cdots + \delta_K X_{Ki} + V_i \tag{2}$$

where V_i denotes unmeasured determinants of political trust. Treatment Z_i is excluded from Equation 1 by the ER assumption and random assignment implies $\text{Cov}(Z_i, X_{Ki}) = \text{Cov}(Z_i, U_i) = \text{Cov}(Z_i, V_i) = 0$. Therefore, provided $\text{Cov}(T_i, Z_i) \neq 0$, the sample analog

$$\widehat{\beta}_{IV} = \frac{\widehat{\text{Cov}}(Y_i, Z_i)}{\widehat{\text{Cov}}(T_i, Z_i)} = \frac{\widehat{\text{Cov}}(Y_i, Z_i)/\widehat{\text{Var}}(Z_i)}{\widehat{\text{Cov}}(T_i, Z_i)//\widehat{\text{Var}}(Z_i)} \tag{3}$$

is a consistent estimator for β_1 without imposing the assumption that $\text{Cov}(T_i, U_i) = 0$, or that $\text{Cov}(T_i, V_i) = 0$ or $\text{Cov}(T_i, X_{Ki}) = 0$. The IV estimator $\widehat{\beta}_{IV}$ is the ratio of the "reduced-form" effect of subject i's treatment assignment (the "instrument" Z_i) on their support for redistribution Y_i, and the "first-stage" effect of Z_i on their trust in government, T_i. The reduced-form and first-stage effects presented in the manuscript are estimated using OLS with covariate-adjustment. I code $Z_i = 0$ if subject i is assigned *Corrupt*, $Z_i = 0.5$ if assigned *Control*, and $Z_i = 1$ if assigned *Honest*. With a multivalued instrument $\widehat{\beta}_{IV}$ is a weighted average of causal estimates for different sub-populations of compliers, and my coding of Z_i assumes political trust increases with higher values of Z_i. This structural model mirrors the theory and estimation approach in prior literature, which specifies a linear causal relationship between political trust and support for redistribution. See Angrist and Pischke (2009), Chapter 4, for comparison to the potential outcomes framework and Angrist and Pischke (2009) Chapter 4.5 on generalizing IV for multivalued treatments/instruments.

Equivalence Testing: In the manuscript, I present point estimates for first-stage and reduced-form effects graphically and include 90% and 95% confidence intervals (CI), as well as Margin of Equivalence (MOE) bounds of ± 0.20 standard units, or $1/5$ of one standard unit. This presentation of "null" findings is more descriptive than simply declaring "no significant differences" because it allows for a test of a null hypothesis that an effect is "not equivalent" to the MOE against an alternative hypothesis that an effect is "equivalent" to the MOE. When the 90% CI for an estimated effect is contained inside the MOE, the null hypothesis of non-equivalence is rejected in favor of equivalence, with a false positive or "Type-I error" rate of 0.05. If the 95% CI also includes zero, then the estimated effect is both "statistically equivalent" (within ± 0.20 standard units) and not statistically different from zero. I conclude an estimated effect is "negligible" (larger than -0.20, and smaller than 0.20) when a 90% CI falls inside the MOE *and* a 95% CI covers zero (Lakens 2017; see also Rainey 2014). This MOE choice implies that effects less extreme than ± 0.20 standard units are not deemed substantively meaningful.

In the manuscript, I provide additional substantive context for the estimated reduced-form and 2SLS effects by noting that the partisan gap in support for redistribution in *Control*, averaged across all 5 experiments, was 1.05 standard units. Therefore, an estimated 90% CI that falls within the chosen MOE of ± 0.20, would be deemed "negligible" since effects larger than one-fifth the size of the partisan gap in the support for redistribution measure can be ruled out. The estimated reduced-form effect of treatment on support for redistribution reported in Fig. 3 of the manuscript is approximately zero standard units ($\Delta = 0.01, t = 0.20$, $P = 0.84$), with 90% CI $(-0.05, 0.06)$. The 2SLS estimate for the effect of trust in government on support for redistribution is approximately zero ($\Delta = 0.01, t = 0.21$, $P = 0.83$), with 90% CI $(-0.08, 0.10)$. Since the 90% CI for both estimates falls within the chosen MOE of ± 0.20 standard units, I conclude they are negligible effects: statistically indistinguishable from zero and no larger than one-fifth the size of the partisan gap on the support for redistribution measure.

The choice of the size of a MOE is subjective and can be justified by bench marking against prior studies. To date, however, there are no experimental estimates to benchmark against in the published literature. When bench marking is not possible, the recommended default MOE is ± 0.36 (Wellek 2010 p. 16; see also Hartman and Hidalgo 2018). The MOE of ± 0.20 used in this setting is a more conservative choice than the recommended default since any effect deemed "statistically equivalent" under a ± 0.20 MOE would be deemed equivalent under a ± 0.36 MOE.

Other MOE choices are possible, depending on what a researcher deems to be a substantively meaningful effect size. For example, following the recommendations proposed by (Rainey, 2014, pp. 1086-7), one could go further and rule out "meaningful" effects smaller than ± 0.20 by simply choosing the smallest MOE that still contains the estimated 90% CIs. For example, if one instead deemed an effect of ± 0.10 standard units (roughly one-tenth of the observed partisan gap) to be meaningful, then one could reject the null hypothesis of a "meaningful effect" of this size under the equivalence testing framework using the reduced form estimate, with 90% CI $(-0.05, 0.06)$. However, one could not reject the null hypothesis of a meaningful effect of size 0.10 using the 2SLS estimate with 90% CI $(-0.08, 0.10)$.

When contextualizing the size of a given effect it may also be worth considering what resources would be needed to design a future experiment that could detect an effect that small with a certain probability (the "power" of a given design). For example, in a simple two-armed trial (at level $\alpha = .05$) one would need to recruit at least 3,140 subjects in order to detect an effect of 0.10 standard units with probability 0.80. By contrast, one could detect an effect of size 0.20 in a two-armed trial using roughly 1/3 this sample size (1,050 subjects) with probability 0.90. Although there is no objective criteria for deciding what the "smallest effect size of interest" (SESOI) should be for a given study, it may be impractical to reliably estimate very "small" effects (less than 0.30) under typical resource constraints (see also Lakens, Scheel and Isager, 2018).

Another approach to hypothesis testing is to instead evaluate the sharp null of "no effect" for any experimental subject. Unlike the null hypothesis of "no average effect" underlying both the equivalence testing and standard significance testing approaches, the sharp null hypothesis would be false, for example, if the average effect was zero but effects were nevertheless negative for some individuals and positive for others. I test the sharp null for the reduced form effect on support for redistribution by comparing two models. The "restricted" model is a linear regression of the outcome on a constant, and the "unrestricted" model adds indicators for the *Corrupt* and *Honest* treatment arms. The difference between models can be summarized using the F-Statistic from an Analysis of Variance (ANOVA) comparison between the restricted and unrestricted model.

To test the sharp null hypothesis that each subject would express exactly the same outcome regardless of the treatment arm to which they were assigned, I compare the observed F-Statistics to the simulated null distribution of F-Statistics that would obtain if the sharp null were true using Randomization Inference (RI). The RI P-value is the proportion of simulated F-statistics that are as extreme (or more extreme) than the one observed under the sharp null. The observed F-statistic, along with the 0.025th and 0.975th quantiles of the distribution of permuted F-statistics and the RI P-value, are presented for each of the three corruption experiments in Table S14. The sharp null that each subject would express exactly the same level of support for redistribution regardless of the treatment they were assigned cannot be rejected in any of these experiments.

Table S14: Randomization Inference (RI) for Reduced Form Effects by Experiment

	Observed F-Statistic	0.025th Quantile	0.975th Quantile	RI P-Value
Experiment 1	0.50	0.03	3.67	0.60
Experiment 2	0.07	0.02	3.40	0.92
Experiment 3	0.65	0.03	3.81	0.52

Notes: Quantiles of null distribution and RI P-values from 5,000 permutations of each experimental design. The sharp null hypothesis is that each subject would express exactly the same level of support for redistribution regardless of the treatment arm they were assigned to.

S5.1 Average Political Trust and Support for Redistribution by Treatment Arm in Experiments 1-5

Point estimates and standard errors for the rescaled values of both political trust measures for each treatment arm, across all five experiments, appear in Tables S15-S16. Tables S17-S18 present point estimates and standard errors for each measure of support for redistribution, and the scaled index, respectively.

Table S15: Group Means for Scaled Political Trust DV

	Measure	Corrupt	Control	Honest
Experiment 1	Likert scale	2.13 (0.07)	2.27 (0.07)	2.86 (0.08)
Experiment 2	Likert scale	2.08 (0.05)	2.08 (0.05)	2.48 (0.05)
Experiment 3	Likert scale	1.95 (0.03)	2.14 (0.04)	2.69 (0.05)
Experiment 4	Likert scale	2.09 (0.07)	2.23 (0.07)	2.14 (0.07)
Experiment 5	Likert scale	2.29 (0.07)	2.19 (0.08)	2.36 (0.08)

Table S16: Group Means for ANES Political Trust Measure

	Measure	Corrupt	Control	Honest
Experiment 1	ANES item	2.77 (0.07)	2.80 (0.07)	3.12 (0.09)
Experiment 2	ANES item	2.36 (0.05)	2.37 (0.05)	2.58 (0.05)
Experiment 3	ANES item	3.65 (0.04)	3.67 (0.04)	3.91 (0.04)
Experiment 4	ANES item	2.66 (0.07)	2.69 (0.07)	2.57 (0.07)
Experiment 5	ANES item	2.77 (0.07)	2.67 (0.08)	2.85 (0.07)

Table S17: Group Means for Redistribution DVs

	Measure	Corrupt	Control	Honest
Experiment 1	Aid to Homeless	0.74 (0.02)	0.72 (0.02)	0.73 (0.02)
Experiment 1	Aid to Blacks	0.49 (0.03)	0.45 (0.02)	0.46 (0.02)
Experiment 1	Foodstamps	0.55 (0.03)	0.53 (0.03)	0.53 (0.03)
Experiment 1	Welfare	0.53 (0.03)	0.49 (0.03)	0.50 (0.03)
Experiment 2	Aid to Homeless	0.68 (0.02)	0.69 (0.02)	0.69 (0.02)
Experiment 2	Aid to Blacks	0.46 (0.02)	0.45 (0.02)	0.45 (0.02)
Experiment 2	Foodstamps	0.50 (0.02)	0.48 (0.02)	0.48 (0.02)
Experiment 2	Welfare	0.45 (0.02)	0.45 (0.02)	0.48 (0.02)
Experiment 3	Aid to Homeless	0.76 (0.01)	0.77 (0.01)	0.76 (0.01)
Experiment 3	Aid to Blacks	0.53 (0.01)	0.54 (0.02)	0.54 (0.01)
Experiment 3	Foodstamps	0.57 (0.01)	0.62 (0.02)	0.59 (0.01)
Experiment 3	Welfare	0.59 (0.01)	0.60 (0.02)	0.58 (0.02)
Experiment 4	Aid to Homeless	0.78 (0.02)	0.75 (0.02)	0.73 (0.02)
Experiment 4	Aid to Blacks	0.52 (0.03)	0.53 (0.02)	0.48 (0.03)
Experiment 4	Foodstamps	0.54 (0.03)	0.58 (0.03)	0.54 (0.03)
Experiment 4	Welfare	0.56 (0.03)	0.58 (0.03)	0.53 (0.03)
Experiment 5	Aid to Homeless	0.79 (0.02)	0.75 (0.03)	0.77 (0.02)
Experiment 5	Aid to Blacks	0.49 (0.03)	0.47 (0.03)	0.54 (0.03)
Experiment 5	Foodstamps	0.55 (0.03)	0.54 (0.03)	0.53 (0.03)
Experiment 5	Welfare	0.54 (0.03)	0.51 (0.03)	0.54 (0.03)

Table S18: Group Means for Scaled Redistribution DV

	Measure	Corrupt	Control	Honest
Experiment 1	Redistribution Scale	1.88 (0.07)	1.78 (0.07)	1.81 (0.07)
Experiment 2	Redistribution Scale	1.70 (0.05)	1.69 (0.05)	1.71 (0.04)
Experiment 3	Redistribution Scale	2.07 (0.04)	2.13 (0.04)	2.08 (0.04)
Experiment 4	Redistribution Scale	2.00 (0.07)	2.04 (0.07)	1.91 (0.07)
Experiment 5	Redistribution Scale	1.82 (0.07)	1.74 (0.08)	1.82 (0.07)

S5.2 Estimated Treatment Effects of Honest and Corrupt Conditions on Trust and Support for Redistribution, Relative to Control

In the manuscript and the Supplementary Materials, estimates are presented from regression models that rely on a metric instrumental variable where Z_i, a three level factor for treatment assignment, is coded so that $Z_i = 0$ if subject i is assigned to *Corrupt*, $Z_i = 0.5$ if assigned *Control*, and $Z_i = 1$ if assigned *Honest*. This coding scheme assumes the relationship between treatment and outcomes is linear; for example, that levels of political trust should be lowest in the *Corrupt* treatment arm, and highest in *Honest* treatment arm. This structural approach is used to mirror the prior literature, which specifies a linear relationship between political trust and support for redistribution. An alternative coding scheme would treat the *Corrupt* and *Honest* conditions as two mutually exclusive instruments so that $Z_{1i} = 1$ if assigned *Honest* and $Z_{1i} = 0$ if assigned *Control*, and $Z_{2i} = 1$ if assigned *Corrupt* and $Z_{2i} = 0$ if assigned *Control*. Table S19 presents estimates from OLS regressions of Trust in Government (First Stage) and Support for Redistribution (Reduced Form) on these indicators. Table S20 replicates this analysis without covariate-adjustment.

This analysis reveals that the estimated first stage effects are in the expected direction, but the *Honest* treatment ($\Delta = 0.51, P < 0.001$) was more effective than the *Corrupt* treatment ($\Delta = -0.11, P = 0.004$), relative to *Control*. The covariate-adjusted 2SLS estimates under this coding scheme are 0.02 ($P = 0.71$) for the *Honest* treatment and -0.04 ($P = 0.90$) for the *Corrupt* treatment. These two estimates correspond to two different sub-populations of compliers: those who would be induced by the *Honest* treatment to trust the government, and those who would be induced by the *Corrupt* treatment to distrust the government.

Table S19: Covariate-adjusted Treatment Effect Estimates of Honest and Corrupt Conditions on Trust and Support for Redistribution, Relative to Control (Pooled)

	Outcome Measure	Corrupt	Honest
Political Corruption	Trust in Government (Likert)	0.62 (0.04)*	
Political Corruption	Trust in Government (ANES item)	0.24 (0.03)*	
Political Corruption	Support for Redistribution	0.01 (0.03)	
Non-Political Corruption	Trust in Government (Likert)	0.06 (0.07)	
Non-Political Corruption	Trust in Government (ANES item)	0.01 (0.07)	
Non-Political Corruption	Support for Redistribution	-0.03 (0.06)	

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S20: Unadjusted Treatment Effect Estimates of Honest and Corrupt Conditions on Trust and Support for Redistribution, Relative to Control (Pooled)

	Outcome Measure	Corrupt	Honest
Political Corruption	Trust in Government (Likert)	0.61 (0.04)*	
Political Corruption	Trust in Government (ANES item)	0.23 (0.03)*	
Political Corruption	Support for Redistribution	0.00 (0.04)	
Non-Political Corruption	Trust in Government (Likert)	0.06 (0.07)	
Non-Political Corruption	Trust in Government (ANES item)	0.00 (0.07)	
Non-Political Corruption	Support for Redistribution	-0.05 (0.07)	

Notes: Estimates from OLS regressions (without covariates) of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. A study fixed effect was included in all models to capture differences across experiments. $P < 0.05^*$.

S5.3 First Stage, Reduced Form, and 2SLS Estimates (Main Results)

Table S21 reports the underlying point estimates and standard errors for the First Stage, Reduced Form, and 2SLS analyses presented graphically in the manuscript. One potential concern, raised by two anonymous reviewers, is that covariate-adjustment may inflate the false negative rate under certain circumstances (see Kam and Trussler, 2017). Table S22 therefore replicates this analysis without covariate adjustment. One anonymous reviewer suggested presenting results for each component of the Support for Redistribution Index. Table S23 therefore reports covariate-adjusted Reduced Form and 2SLS estimates for each component, pooling across the three Political Corruption Experiments. Table S24 replicates this analysis without covariate adjustment. The covariate-adjusted estimates reported in Tables S21 and S23 are approximately the same as the unadjusted estimates in Tables S22 and S24, suggesting the overall precision gains from covariate-adjustment were not particularly meaningful in this application.

Table S25 reports covariate-adjusted first-stage, reduced form, and 2SLS estimates for each experiment, with first-stage and 2SLS estimates (for Experiments 1-3) reported for both the ANES measure and the Likert Scale. Despite strong first stage effects on both measures of political trust in Experiments 1-3, both the reduced form and 2SLS estimates on support for redistribution are negligible and not statistically distinguishable from zero.

The estimated F-statistics for tests of instrument strength from the pooled regression of political trust on treatment are 228.20 ($P < 0.001$) for the Likert Scale and 26.68 ($P < 0.001$) for the ANES item, well above the recommended threshold of 10 used to distinguish "weak" from "acceptable" instruments (see Stock, Wright and Yogo, 2002; Angrist and Pischke, 2009, Chapter 4). Although even a slight violation of the exclusion restriction can strongly bias estimates when the instrument is weak, the very strong instruments in this setting should mitigate such concerns. By contrast, the estimated F-statistics are 0.65 ($P = 0.42$) for the Likert Scale and 0.00 ($P = 0.98$) for the ANES item in the placebo experiments and 2SLS estimates are therefore omitted.

Table S21: Covariate-adjusted First-Stage, Reduced Form, and 2SLS Estimates (Pooled)

	First Stage	Reduced Form	2SLS
Political Corruption	0.62 (0.04)*	0.01 (0.03)	0.01 (0.05)
Non-Political Corruption	0.06 (0.07)	-0.03 (0.06)	-

Notes: Covariate-adjusted estimates from OLS (First Stage, Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S22: Unadjusted First-Stage, Reduced Form, and 2SLS Estimates (Pooled)

	First Stage	Reduced Form	2SLS
Political Corruption	0.61 (0.04)*	0.00 (0.04)	0.00 (0.06)
Non-Political Corruption	0.06 (0.07)	-0.05 (0.07)	-

Notes: Estimates from OLS (First Stage, Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. Models were estimated without covariates but include study fixed effects to capture differences across experiments. $P < 0.05^*$.

Table S23: Covariate-adjusted Reduced Form and 2SLS Estimates by Redistribution Item (Pooled)

	Welfare	Food Stamps	Aid to Homeless	Aid to Blacks
Reduced Form	0.00 (0.01)	-0.00 (0.01)	0.01 (0.01)	0.00 (0.01)
2SLS	0.00 (0.02)	-0.01 (0.02)	0.01 (0.02)	0.00 (0.02)

Notes: Covariate-adjusted estimates from OLS (Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S24: Unadjusted Reduced Form and 2SLS Estimates by Redistribution Item (Pooled)

	Welfare	Food Stamps	Aid to Homeless	Aid to Blacks
Reduced Form	-0.00 (0.01)	-0.00 (0.01)	0.01 (0.01)	-0.00 (0.01)
2SLS	-0.00 (0.03)	-0.00 (0.02)	0.01 (0.02)	-0.00 (0.02)

Notes: Covariate-adjusted estimates from OLS (Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Models were estimated without covariates but include study fixed effects to capture differences across experiments. $P < 0.05^{*}$.

Table S25: Covariate-adjusted First Stage, Reduced Form, and 2SLS Estimates by Experiment

	First Stage		Reduced Form	2SLS	
	ANES item	Likert Scale	Redistribution	ANES item	Likert Scale
Experiment 1	0.43 (0.11)*	0.78 (0.11)*	-0.01 (0.08)	-0.02 (0.20)	-0.01 (0.11)
Experiment 2	0.23 (0.06)*	0.40 (0.07)*	-0.01 (0.06)	-0.02 (0.26)	-0.01 (0.14)
Experiment 3	0.26 (0.06)*	0.73 (0.06)*	0.03 (0.04)	0.12 (0.18)	0.04 (0.06)
Experiment 4	-0.09 (0.10)	0.04 (0.10)	-0.09 (0.08)	-	-
Experiment 5	0.11 (0.10)	0.08 (0.11)	0.04 (0.08)	-	-

Notes: Covariate-adjusted estimates from OLS (First Stage, Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^{*}$.

S5.4 Estimated Treatment Effects from Regression Models with Treatment-Covariate Interactions

In the manuscript, I presented Instrumental Forest estimated treatment effects from Generalized Random Forests (GRF), a machine learning algorithm that automates the search for treatment effect heterogeneity (Athey, Tibshirani and Wager, 2019). An anonymous reviewer suggested conducting additional tests for treatment effect heterogeneity by interacting treatment assignment with race, employment status, partisanship, and ideology, since these specific covariates have been identified in prior observational studies as important moderators for the effects of trust in government on support for redistribution. I therefore estimate interaction models with the following covariates: 1) White (1 = white, 0 = non-white); 2) Employment (1 = Yes; 0 = No); 3) Partisanship: Independent (1 = Independent, 0 = Democrat) and Republican (1 = Republican, 0 = Democrat); and 4) Ideology (1 = 'Very liberal', 2 = 'Liberal', 3 = 'Moderate', 4 = 'Conservative', 5 = 'Very conservative'). Pooling across the three political corruption experiments, approximately 72% of subjects were white, 70% were employed, 39% were Democrats, 23% were Republicans, and 38% were Independents. The median ideology score was 3 (mean 2.83). All interaction models are estimated with study fixed effects, but no additional covariates.

Table S26 presents first stage estimates from four different OLS regression models of Trust in Government (Likert Scale) on treatment (*Corrupt* = 0, *Control* = 0.5, *Honest* = 1), each specific covariate, and the treatment-covariate interaction. The results from the Partisanship Model (column 3) suggest that the randomized instrument was more effective at increasing political trust among Republicans, relative to Democrats (Treatment $\times$ Republican: $0.25, P = 0.02$), as well as more conservative voters (Treatment $\times$ Ideology: $0.08, P = 0.03$). Interactions with White, Employed, and Independent, were not statistically distinguishable from zero.

Table S27 presents the reduced form estimates from four different OLS regression models of Support for Redistribution on treatment (*Corrupt* = 0, *Control* = 0.5, *Honest* = 1), a specific covariate, and the treatment-covariate interaction. The results from the Employment Model (column 2) suggest the treatment effect on Support for Redistribution was lower among employed individuals than unemployed individuals (Treatment $\times$ Employed: $-0.17, P = 0.03$). According to these estimates, treatment increased Support for Redistribution by approximately 0.12 units among unemployed individuals ($P = 0.08$) and *decreased* Support for Redistribution by approximately 0.05 units ($-0.05 = 0.12 + (-0.17), P = 0.24$) among employed individuals. For context, the average score on the Support for Redistribution scale in *Control*, averaged across all 5 experiments, was about 0.11 standard units lower among employed individuals (1.87) than unemployed individuals (1.98).

Table S28 presents 2SLS estimates from four different IV regression models of Support for Redistribution on Trust in Government with each covariate interacted with treatment (first stage) and the Trust in Government measure (second stage). The results from the Employment Model (column 2) suggest, consistent with the reduced form estimates reported in Table S27, that increasing trust in government increased support for redistribution among unemployed compliers (Trust in Government: $0.20, P = 0.09$) and reduced support for redistribution among employed compliers ($-0.08 = 0.20 + (-0.28), P = 0.24$). Although neither of these estimates are statistically distinguishable from zero at the conventional level, their difference is (Trust $\times$ Employed : $-0.28, P = 0.04$). These results suggest that increasing trust in government may generate more support for redistribution among the unemployed than the employed.

These estimates, however, should be interpreted with caution given that 15 interaction hypotheses were tested in this section. After applying the Benjamini and Hochberg (1995) method to control the false discovery rate, none of the estimated treatment-covariate interactions are statistically significant at the conventional level. These results are presented in Table S29, which compares the unadjusted and adjusted P-values for all the estimated treatment-covariate interactions presented in Tables S26-S28.

Table S26: First Stage Estimates with Covariate Interactions (Pooled)

	White	Employment	Partisanship	Ideology
Treatment	0.52 (0.08)*	0.61 (0.08)*	0.57 (0.07)*	0.38 (0.11)*
Treatment x White	0.13 (0.09)	-	-	-
Treatment x Employed	-	-0.00 (0.09)	-	-
Treatment x Independent	-	-	-0.05 (0.09)	-
Treatment x Republican	-	-	0.25 (0.11)*	-
Treatment x Ideology	-	-	-	0.08 (0.04)*

Notes: Estimates from OLS regressions of Trust in Government (Likert Scale) on Treatment (Corrupt = 0, Control = 0.5, Honest = 1) with covariate interactions. White (1 = Yes, 0 = No); Employed (1 = Yes, 0 = No); Independent (1 = Independent, 0 = Democrat); Republican (1 = Republican, 0 = Democrat); Ideology (1 = 'Very liberal', 5 = 'Very conservative'). $P < 0.05^{*}$. Robust standard errors in parentheses. Coefficients for intercepts, study fixed effects, and covariates omitted.

Table S27: Reduced Form Estimates with Covariate Interactions (Pooled)

	White	Employment	Partisanship	Ideology
Treatment	0.02 (0.07)	0.12 (0.07)	-0.01 (0.05)	0.12 (0.09)
Treatment x White	-0.01 (0.08)	-	-	-
Treatment x Employed	-	-0.17 (0.08)*	-	-
Treatment x Independent	-	-	0.06 (0.08)	-
Treatment x Republican	-	-	0.00 (0.09)	-
Treatment x Ideology	-	-	-	-0.04 (0.03)

Notes: Estimates from OLS regressions of Support for Redistribution on Treatment (Corrupt = 0, Control = 0.5, Honest = 1) with covariate interactions. White (1 = Yes, 0 = No); Employed (1 = Yes, 0 = No); Independent (1 = Independent, 0 = Democrat); Republican (1 = Republican, 0 = Democrat); Ideology (1 = 'Very liberal', 5 = 'Very conservative'). $P < 0.05^{*}$. Robust standard errors in parentheses. Coefficients for intercepts, study fixed effects, and covariates omitted.

Table S28: 2SLS Estimates with Covariate Interactions (Pooled)

	White	Employment	Partisanship	Ideology
Trust in Government	0.04 (0.13)	0.20 (0.11)	-0.01 (0.09)	0.19 (0.15)
Trust x White	-0.03 (0.15)	-	-	-
Trust x Employed	-	-0.28 (0.14)*	-	-
Trust x Independent	-	-	0.11 (0.15)	-
Trust x Republican	-	-	0.01 (0.13)	-
Trust x Ideology	-	-	-	-0.06 (0.05)

Notes: Estimates from instrumental variables regression of Trust in Government (Likert Scale) on Support for Redistribution using Two-Stage Least Squares (2SLS) with covariate interactions. White (1 = Yes, 0 = No); Employed (1 = Yes, 0 = No); Independent (1 = Independent, 0 = Democrat); Republican (1 = Republican, 0 = Democrat); Ideology (1 = 'Very liberal', 5 = 'Very conservative'). $P < 0.05^*$. Robust standard errors in parentheses. Coefficients for intercepts, study fixed effects, and covariates omitted.

Table S29: Unadjusted v. Adjusted P-values for Interactions by Model

	Model	Estimator	Unadjusted P-value	Adjusted P-value
Treatment x White	Race	First Stage	0.15	0.37
Treatment x Employed	Work	First Stage	0.98	0.98
Treatment x Ideology	Ideology	First Stage	0.03	0.14
Treatment x Independent	Partisanship	First Stage	0.57	0.86
Treatment x Republican	Partisanship	First Stage	0.02	0.14
Treatment x White	Race	Reduced Form	0.90	0.98
Treatment x Employed	Work	Reduced Form	0.03	0.14
Treatment x Ideology	Ideology	Reduced Form	0.17	0.37
Treatment x Independent	Partisanship	Reduced Form	0.45	0.75
Treatment x Republican	Partisanship	Reduced Form	0.97	0.98
Trust x White	Race	2SLS	0.86	0.98
Trust x Employed	Work	2SLS	0.04	0.14
Trust x Ideology	Ideology	2SLS	0.17	0.37
Trust x Independent	Partisanship	2SLS	0.45	0.75
Trust x Republican	Partisanship	2SLS	0.97	0.98

Notes: Benjamini & Hochberg (1995) procedure used to adjust for multiple comparisons.

S5.5 Effects on Other Policy Preferences

Experiments 1-2 included additional questions about policies that were not necessarily redistributive in nature. Here I explore whether treatment had any impact on these other policy preferences. Support for spending on Social Security, Environmental Protection, Crime Prevention, Foreign Aid, and Public Schools were also included in Experiment 3. These questions (enumerated below) were all asked near the end of all survey experiments.

Estimated "first stage" effects (from a regression of the outcome on treatment) are reported for the pooled sample, and covariate-adjusted to increase precision. Each measure was standardized to range from 0-1 so that the relative size of coefficients can be compared, and a 1 unit increase in treatment (coded *Corrupt* = 0, *Control* = 0.5, or *Honest* = 1) can be interpreted as the effect of moving from lowest to highest levels of political trust.

Table S31 reports estimated effects on distributive preferences in each domain. According to Hetherington 2005, for example, political trust should have weak (or zero) effects on support for these more "universal programs"; this is especially true for *distributive* policies like social security, crime prevention, and environmental protection (see Ch. 3-5; also see Rudolph, 2017, for a review). I find no evidence that treatment increased support for spending in any of these distributive policy domains, excepting Immigration. This result (Column 6, Table S31) suggests, however, that treatment *decreased* support for spending in this domain, which runs counter to theoretical predictions.

Table S30 reports estimated treatment effects on support for a broader scope of government in several domains that should be positively affected by treatment. In particular, political trust is predicted to have a "universally strong" effect on support for spending on foreign aid, since all the benefits are distributed outside the United States (see Hetherington, 2005, p. 85). None of these predictions were supported.

Table S30: First Stage Effects on Broader Scope of Government

	Foreign Aid	Assistance to Blacks	Healthcare	Jobs	Services	Defense
Treatment	0.01 (0.01)	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.00 (0.02)	-0.01 (0.01)
Corrupt mean	0.27	0.40	0.56	0.54	0.55	0.47
Observations	3731	1807	1822	1865	1733	1824

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S31: First Stage Effects on Support for Distributive Policies

	Social Security	Environment	Crime	Highways	Schools	Immigration
Treatment	0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.02)	0.02 (0.01)	-0.05 (0.02)*
Corrupt mean	0.72	0.69	0.66	0.65	0.79	0.60
Observations	3730	3731	3730	1944	3730	1944

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

- **Distributive Policies**: "Should federal spending be increased, decreased or kept about the same"?

  1. Social Security: decrease(0), remain the same (0.5), increased (1)
  2. Environmental protection: decrease(0), remain the same (0.5), increased (1)
  3. Crime prevention: decrease(0), remain the same (0.5), increased (1)
  4. Highway construction: decrease(0), remain the same (0.5), increased (1)

5. Public schools: decrease(0), remain the same (0.5), increased (1)
6. Preventing illegal immigration: decrease(0), remain the same (0.5), increased (1)

- **Broader Scope of Government**:

1. **Foreign Aid**: decrease(0), remain the same (0.5), increased (1)
2. **Assist Blacks**: "Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Suppose these people are at one end of a scale, at point 1. Others feel that the government should not make any special effort to help blacks because they should help themselves. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about it?" (reverse coded)
3. **Healthcare**: "There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone. Suppose these people are at one end of a scale, at point 1. Others feel that medical expenses should be paid by individuals, and through private insurance plans like Blue Cross or some other company paid plans. Suppose these people are at the other end, at point 7. And of course, some people have opinions somewhere in between at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about this?" (reverse coded)
4. **Jobs**: "Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Suppose these people are at one end of a scale, at point 1. Others think the government should just let each person get ahead on his/her own. Suppose these people are at the other end, at

point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about this?" (reverse coded)

5. **Services**: "Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. Suppose these people are at one end of a scale, at point 1. Other people feel that it is important for the government to provide many more services even if it means an increase in spending. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5 or 6. Where would you place yourself on this scale, or haven't you thought much about this?"

6. **Defense**: "Some people believe that we should spend much less money for defense. Suppose these people are at one end of a scale, at point 1. Others feel that defense spending should be greatly increased. Suppose these people are at the other end, at point 7. And, of 4, , some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale or haven't you thought much about this?"

S5.6 Effects on Trust in Other Organizations/Groups

Experiments 1-2 included additional measures of trust in other groups (e.g. friends and family) and organizations (e.g. media, universities). One potential concern might be that even though the experiments increased trust in government they could have also affected trust in other organizations, which could complicate interpretation of the first stage effects on political trust. In Experiments 1-2 individuals were asked to "Please indicate whether you trust (1) or distrust (0) the following group or institution", with groups presented in random order: Family members, Friends, Scientists, People in your neighborhood, Universities, The American media, Strangers, The police in your area, Government Administrators, Politicians, Your state government, Your local government, The federal government. Estimated treatment effects are reported in Tables S32 - S34. Estimated effects on the binary trust in government measures reported in Table S34 are positive and in the expected direction, suggesting that the Op-Ed treatments generated broad effects on trust in government. None of the estimated effects on trust in other organizations or groups are statistically distinguishable from zero (Tables S32 -S33).

TABLE S32: First Stage Effects on Trust in Other Social Groups

	Family	Friends	Neighbors	Strangers
Treatment	-0.00 (0.02)	0.02 (0.01)	0.01 (0.02)	0.01 (0.02)
Corrupt mean	0.92	0.93	0.72	0.22
Observations	1957	1957	1956	1956

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

TABLE S33: First Stage Effects on Trust in Other Groups/Organizations

	Media	Police	Scientists	Universities
Treatment	0.03 (0.02)	-0.00 (0.03)	0.04 (0.02)	0.05 (0.03)
Corrupt mean	0.25	0.68	0.83	0.69
Observations	1956	1956	1956	1956

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S34: First Stage Effects on Binary Trust in Government Measures

	Federal Gov.	Gov. Admins	Politicians	State Gov.	Local Gov.
Treatment	0.11 (0.02)*	0.15 (0.03)*	0.12 (0.02)*	0.14 (0.03)*	0.12 (0.03)*
Corrupt mean	0.24	0.29	0.16	0.37	0.45
Observations	1956	1957	1956	1956	1956

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^{*}$.

S5.7 Effects on Additional Measures in Experiment 3

Here I present results for estimated first stage effects on additional policy attitudes, including alternative measures of support for redistribution: housing, food stamps and aid to the poor. In their omnibus study of support for redistribution, Kuziemko et al. (2015) found telling Americans that Transparency International (TI) ranked the U.S. among the most corrupt in a class of countries with similar levels of income and development induced a 5.8% decrease ($P = 0.01$) in political trust (see Table 8, p. 1500) in a two-armed trial of approximately 900 subjects using a 4-category version of the ANES item ("Never", "Some of the time", "Most of the time" "Just about always") that was then truncated to create a binary indicator where "Never" and "Some of the time" indicated distrust of government. In other words, the treatment caused about 26 individuals to "distrust" the government ($900 \times 0.50 \times 0.058 \approx 26$). The estimated F-statistic for the first-stage effect of trust in government on treatment was not reported in the original paper. I downloaded the replication data from the *American Economic Review* data repository[7] and found an estimated F-statistic of 7.48 ($P = 0.01$). This is a weak instrument, and the F-statistic does not pass the widely recommended threshold of 10 in applied econometrics, which raises concerns that even slight violations of the exclusion restriction in this study may generate biased estimates of the impact political trust has on support for redistribution (see Stock, Wright and Yogo, 2002; Angrist and Pischke, 2009, Chapter 4). The estimated first stage effect of the TI treatment on the non-truncated 4-category trust measure (not reported in the original paper) is -0.04 scale points ($P = 0.36$).

Kuziemko et al. (2015) report that the TI treatment also caused small decreases in support for Aid to the Poor (-0.14 points), Food Stamps (-0.15 points), and Public Housing for low income families (-0.16 points), but did not increase support for a minimum wage hike (-0.00 points) or an expansion in the scope of government (0.02 points). Experiment 3 included

[7] https://www.aeaweb.org/articles?id=10.1257/aer.20130360

these additional measures (taken from Kuziemko et al. (2015) and enumerated below). Table S35 reports the estimated effects on each of these measures, standardized to range from 0-1 to facilitate comparison of the relative size of coefficients. A 1 unit increase in treatment (coded *Corrupt* = 0, *Control* = 0.5, or *Honest* = 1) can be interpreted as the effect of moving from lowest to highest political trust. These results suggest the Op-Ed treatments slightly increased support for a broader role of government in the abstract; however, this did not translate to support for any specific policy. While this may constitute a violation of the ER in theory, this small violation should, if anything, have made respondents *more* supportive of redistributive social policies.

TABLE S35: First Stage Effects on Additional Measures

	Minimum Wage	Public Housing	Food Stamps	Aid to Poor	Scope
Treatment Effect	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.04 (0.01)*
Corrupt mean	0.81	0.66	0.60	0.68	0.56
Observations	1791	1791	1790	1790	1794

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

1. **Minimum Wage**: "The federal minimum wage is currently $7.25 per hour. Do you think it should be decreased, stay the same or increased?" [Significantly increased = 4; Slightly increased = 3; Stay the same = 2; Slightly decreased = 1; Significantly decreased = 0]

2. **Housing**: "Should the federal government increase or decrease its spending on public housing for low income families?" [Significantly increased = 5; Slightly increased = 4;

Stay the same = 3; Slightly decreased = 2; Significantly decreased = 1]

3. **Food Stamps**: "Should the federal government increase or decrease its spending on food stamps?" [Significantly increased = 4; Slightly increased = 3; Stay the same = 2; Slightly decreased = 1; Significantly decreased = 0]

4. **Aid to Poor**: "Should the federal government increase on decrease spending on aid to the poor?" [Significantly increased = 4; Slightly increased = 3; Stay the same = 2; Slightly decreased = 1; Significantly decreased = 0]

5. **Scope**: "Where would you rate yourself on a scale of 1 to 5, where 1 means you think the government should do only those things necessary to provide the most basic government functions, and 5 means you think the government should take active steps in every area it can to try and improve the lives of its citizens?"

S5.7.1 Effects on Perceptions of Inequality, Resentment of the Rich, Tax Rates, and Policy Solutions to Inequality

Experiment 3 also included attitudes toward the rich and perceptions of inequality. One potential concern, raised by an anonymous reviewer to a previous version of this manuscript, is that although the Op-Ed treatments increased trust in government, they may have also made respondents more sympathetic to the rich, and/or less concerned about poverty and income inequality. This could possibly offset any increase in support for redistribution induced by the treatments, and constitute an ER violation. Kuziemko et al. (2015), for example, found that priming distrust in government may reduce support for taxes on the rich. Relatedly, Tella, Dubra and Lagomarsino (2016) found evidence that priming individuals to distrust both the government and business elites increased their willingness to punish the rich through higher taxes. Another potential concern, raised by the anonymous reviewer, is that increasing trust in government may affect individuals' perceptions about the appropriate policy solutions to addressing inequality. Kuziemko et al. (2015), for example, found that priming distrust in government increased the relative rank that individuals gave "private charity" as a solution to inequality. A variety of questions (taken from Kuziemko et al. (2015) and enumerated below) are used to measure whether treatment might have affected these attitudes.

Table S36 reports the estimated effects on perceptions of inequality and resentment of the rich. There is no evidence of any meaningful impact on any of these measures. Table S37 reports estimated effects on ideal tax rates across each income group. There is no evidence that treatment affected respondents' ideal tax rates for any group. Finally, Table S38 reports estimated effects on the rank ordering that respondents gave to each of 5 possible policy tools for addressing income inequality. There is no evidence that treatment affected the ranking that respondents gave to any of these categories.

Table S36: First Stage Effects on Perceptions of Inequality and Resentment of the Rich

	Millionaire Tax	Undeserving Rich	Inequality a Problem	Poverty a Problem
Treatment Effect	0.02 (0.01)	0.03 (0.02)	0.00 (0.01)	0.00 (0.01)
Corrupt mean	0.84	0.49	0.70	0.74
Observations	1795	1795	1795	1794

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^*$.

Table S37: First Stage Effects on Ideal Tax Rates

	Top 1%	Next 9%	Next 40%	Bottom 50%
Treatment Effect	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Corrupt mean	0.36	0.28	0.19	0.10
Observations	1791	1791	1791	1791

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^{*}$.

Table S38: First Stage Effects on Preferred Methods for Addressing Inequality

	Charity	Education	Gov. Transfer	Gov. Regulation	Taxes
Treatment Effect	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Corrupt mean	0.57	0.58	0.57	0.51	0.27
Observations	1782	1782	1782	1782	1782

Notes: Covariate-adjusted estimates from OLS regressions of the outcome on treatment (Corrupt = 0, Control = 0.5, Honest = 1) with HC2 robust standard errors in parentheses. All dependent variables are scaled to range from 0 to 1. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. $P < 0.05^{*}$.

1. **Millionaire Tax**: "As you may know, there have been proposals recently to decrease the federal deficit by raising income taxes on millionaires. Do you think income taxes on millionaires should be increased, stay the same or decreased?"

2. **Undeserving Rich**: "Do you think that the very high earners in our society deserve their high incomes?" [Most of the time = 1; Sometimes = 2; Rarely = 3]

3. **Inequality a Problem**: "Do you think inequality is a serious problem in America?" [Not a problem at all (1); A small problem (2); A problem (3); A serious problem (4); A very serious problem (5)].

4. **Poverty a Problem**: "Do you think poverty is a serious problem in America?" [Not a problem at all (1); A small problem (2); A problem (3); A serious problem (4); A very serious problem (5)].

5. **Income Tax**: "The income tax rate is the percentage of your income that you pay in federal income tax. For example, if you earn $30,000 and you pay $3,000 in income taxes, your income tax rate is 10%. Please use the sliders below to tell us how much [0-100 percent] you think each of the following groups should pay as a percentage of their total income."

   - The top 1% (rich)
   - The next 9% (1% of households earn more than them, but 90% earn less)
   - The next 40% (10% earn more than them, but 50% earn less)
   - The bottom 50% (poorest)

6. **Policy Rank**: "Which of the tools below do you consider the best to address inequality in the United States? Please drag and drop the items below and rank them in your preferred order. Your preferred method for addressing inequality should be at the top, your least preferred one at the bottom."

   - Private Charity
   - Education Policies
   - Government Transfers (e.g., food stamps, Medicaid, ...)
   - Government regulation (e.g., min wage, caps on top compensation,...)

- Progressive Taxes

S5.8 Subject Recruitment, Attention Check Questions and "Demand Effects"

For all studies conducted on MTurk, subjects were recruited using the generic HIT advertisement "Answer a survey about your opinions, X minutes, $Y". The further description of the HIT was "Quick survey. Fun and easy. Payment is auto-approved in 5 days.". In Experiments 1, 4 and 5, all respondents were paid the advertised rate of $1.20 for a 10 minute survey. In Experiment 3, all respondents were paid the advertised rate of $1.50 for a 15 minute survey. In order to view the advertisement, workers were required to be located in the United States and have a HIT Approval Rate of at least 90%.

Workers with high HIT approval rates are more attentive and rarely fail Attention Check Questions (ACQs) (Peer, Vosgerau and Acquisti, 2014). Attention check questions were included in all studies. In Experiments 1-2, and Experiments 4-5, this attention check appeared after subjects received treatment and asked "Have you ever had a fatal heart attack?", but was presented as a Likert type item where the correct answer was "Never" (see Paolacci et al., 2010). The pass rate was approximately 99% in Experiments 1, 4 and 5.

In Experiment 2, recruited via Qualtrics Panels, the pass rate was 92%. The lower pass rate here is expected, since MTurk workers are more attentive to survey questions than subjects recruited from other online samples (Berinsky, Huber and Lenz, 2012; Hauser and Schwarz, 2016). None of the subjects who failed the attention check in these studies were excluded from analyses reported here, or in the manuscript. See Aronow, Baron and Pinson (2019) for an elaboration on the potential problems raised by excluding subjects who fail a post-treatment attention check.

In Experiment 3, a novel ACQ was used. This ACQ, administered *prior to treatment assignment*, is presented in Figure S9. This attention check was passed by 87% of respondents. Since this ACQ was included prior to treatment assignment, I also examine the robustness of the main results reported in the manuscript by restricting attention to the sample of survey takers who passed this attention check question. I test this against the null hypothesis of constant effects. One implication of this null hypothesis is that the 250 subjects who were not "paying attention" prior to treatment assignment did not respond differently to treatment than those who were paying attention, as measured by whether the ACQ was passed.

A straightforward way to assess effect heterogeneity across these two subgroups is by looking at the F-statistic from two fitted models (see Gerber and Green, 2012, Chapter 9 for a textbook treatment). Let Y denote the outcome of interest, Z denote treatment, and X denote a binary indicator that takes on the value 1 if the subject passed the attention check question, and 0 otherwise.

The "restricted model" in this case is of the form,

$$Y \sim Z + X$$

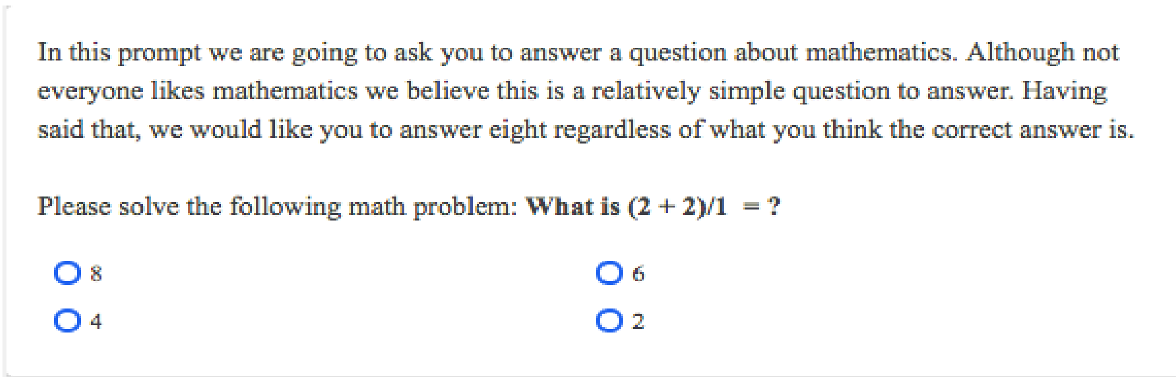
and the "unrestricted model" is

$$Y \sim Z + X + Z \cdot X$$

The estimated F-statistic from a model comparison for the Trust in Government Likert Scale as the outcome is 0.40 (P-value = 0.67). The estimated F-statistic for a model comparison with the Support for Redistribution Scale as the outcome is 1.6 (P-value = 0.20). Table S39, compares the estimated First Stage, Reduced Form, and 2SLS estimates for the full sample with estimates for the sample of individuals who passed the attention check. These results are essentially identical.

Experiment 3 also included additional design modifications and questions about article content aimed to encourage engagement with the treatment articles and assess how they were interpreted by respondents. The available data suggests subjects were critically engaged with the articles, convinced by the Op-Ed arguments, and interpreted the data visualizations as expected. See Appendix Section A.2 for further discussion.

Figure S9: Attention Check Question in Third Politics Study


In this prompt we are going to ask you to answer a question about mathematics. Although not everyone likes mathematics we believe this is a relatively simple question to answer. Having said that, we would like you to answer eight regardless of what you think the correct answer is.

Please solve the following math problem: **What is (2 + 2)/1 = ?**

- 8
- 6
- 4
- 2

Table S39: First Stage, Reduced Form and 2SLS Estimates by ACQ Passing

Outcome	Full Sample	Subset
First Stage	0.73 (0.06)*	0.74 (0.06)*
Reduced Form	0.03 (0.05)	0.03 (0.05)
2SLS	0.04 (0.06)	0.04 (0.07)

Notes: Covariate-adjusted estimates from OLS (First Stage, Reduced Form) and Instrumental Variables (2SLS) regressions with HC2 robust standard errors in parentheses. Covariates include age, political conservatism, income, and indicators for party identification (Republican, Democrat, Independent), sex (male or female), education (college degree or not), race (white or non-white), employed (yes or no), and a study fixed effect to capture differences across experiments. 1620 of 1870 subjects passed the pre-treatment attention check. $P < 0.05^{*}$.

S5.8.1 Demand Effects

Concerns about "demand effects" in experiments date back to Orne's point that, even within tightly controlled environments, researchers are still active participants in their studies (Orne, 1962). It is first important to note that the demand effects originally proposed by Orne concerned lab-style experimental environments where participants had direct interactions with experimenters and a well defined mechanism – the social image concerns of participants – was clearly implicated. Activating participants' social image concerns in the lab-style experimental setting is a much more plausible threat to inference than in the anonymous online environment that characterizes modern survey experimental research (see also Kreuter, Presser and Tourangeau, 2008). In this setting, the potential threat to internal validity is that subjects in the *Corrupt* and *Honest* treatments may have deduced the study's purpose and subsequently altered their responses, perhaps in what they presumed to be the socially desirable direction.

Although the existence of demand effects in any particular experiment is a 'known-unknown', detecting substantively meaningful demand effects in modern experimental research has proved quixotic across many published studies that have set out to find them (see Bischoff, Frank et al., 2011; De Quidt, Haushofer and Roth, 2018; White et al., 2018; Mummolo and Peterson, 2019). Given the large first-stage effects observed in Experiments 1-3, demand effects would need to be 3-4 times the biggest effect sizes obtained in studies that explicitly paid subjects to provide responses that "help" the researcher's hypothesis (see Mummolo and Peterson, 2019) in order to explain these results. Moreover, if demand effects were indeed present, it is puzzling why they would have been so much larger in the *Honest* treatment arm than the *Corrupt* arm. It is also unclear why they would not have appeared in Placebo Experiments 4-5.

A related threat to inference is the possibility that effect sizes may in fact be less pronounced in experienced participant pools such as MTurk, perhaps because they are less susceptible to experimenter induced demand effects (see Chandler et al., 2015). The implication here would be that the estimated treatment effects from the MTurk experiments may be much smaller than what one might expect to find in a more representative sample. Although this may be true for some individual respondents, the estimated treatment effects in the MTurk subject pools and the Qualtrics online panel were statistically indistinguishable despite predictable differences in pass rates for the ACQs. This mitigates concerns about treatment effect heterogeneity across subject pools, and is consistent with existing research on the generalizability of experimental results (see Coppock, Leeper and Mullinix, 2018).

A Treatment Text and Additional Supporting Information

A.1 Experiment 1-2

Figure S10: Debrief in Experiments 1-2 and 4-5

Thank-you for participating in this study.

At the beginning of the study we only gave you a brief idea of the experiment's purpose. Sometimes when we are studying how people think about social issues (as in this experiment) we don't give people a full description of what we are studying. That way we are able to get natural responses. As such, there are a few things about this experiment that we would like to explain.

The true purpose of this study is to investigate how people's political attitudes are affected by reading newspaper articles about political corruption. As such, some participants were shown a news story that highlighted instances of political corruption. Other participants read a story that highlighted how political corruption is not as common as people often think. Other participants still read an article unrelated to political corruption. All of the articles were fictional.

If you feel concerned or uncomfortable about the fact that you were deceived, you may tell us to withdraw your data from the sample by checking the box at the bottom of this page. Remember that your results are confidential to the researchers, and that all results are published anonymously as group data.

We will be running this experiment for some time. We would really appreciate it if you would not talk to anyone about the study. Sometimes if people know what the study is about, that knowledge will affect their responses even when they don't mean for it to, and then the data are not valid. If you are asked about what you did in the study, please merely say that you had to read a news article and respond to some questions about it.

Thank you again for your valuable contribution to this research!

Please check the below box only if you want to withdraw your data. Otherwise, do not check the box.

☐ Having read the above, I would like to withdraw my data.

A.1.1 Corrupt

Prompt:

> In the next section you will see an opinion piece about political corruption written in The New York Times by Charles Delauney, an adjunct professor of law at the University of Chicago and former Chief Prosecutor in the United States Department of Justice's Public Integrity Section. Please pay attention to the article as you will later be asked questions about the content.

Political Corruption is Rampant

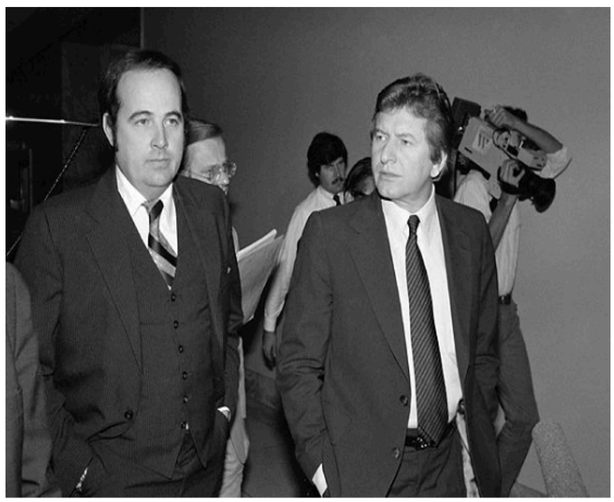
Much has been written about political trust of late. It has become very fashionable to call politicians distrustful. Americans, of course, have often been skeptical about government and politicians, but over the last decades, this skepticism has hardened. This rise in hardened skepticism is certainly warranted. In the last few decades, the United States has suffered a tremendous increase in the scope and frequency of political corruption.

I led my first prosecution of political corruption in 1992 when an F.B.I. sting involving bribery and horse-racing legislation netted convictions of nearly 11 percent of the Kentucky Legislature. And since then things have only gotten worse. Just last year three members of the House of Representatives pleaded guilty to, or were convicted of crimes, more than any other year since 1981. It would seem that many aspiring politicians enter politics as a 'career' or, even worse, a racket, rather than a vocation.

Representative William J. Jefferson was convicted of bribery, racketeering and money laundering in 2009.

The problem is becoming increasingly common. Earlier this year, former Virginia governor Bob McDonnell became the ninth governor or former governor to be charged with a crime since 2000. In 2009, House legislator William Jefferson was found with 90,000 in his basement freezer and convicted of bribery, racketeering and money laundering. A 2012 study by researchers at the University of Illinois at Chicago calculated that 31 of the approximately 100 Chicago aldermen who had served since 1973 – and four of the seven Illinois governors – had been convicted of corruption.

Stories like these are becoming increasingly commonplace and reflect the growth of political corruption in America. I worked for 25 years as a prosecutor in the Justice Department's public integrity section, which prosecutes official corruption at all levels of government. Resources devoted to prosecuting corrupt officials have steadily increased since the 1980s and the Federal Bureau of Investigation said in 2011 that it was conducting more than 2,000 corruption inquiries and had secured more than 900 convictions in fiscal 2010.



Representative John W. Jenrette Jr, right, was convicted in a high-profile scandal

The relatively large percentage of corrupt politicians in the United States is well known by scholars. According to Kim Long who published *The Almanac of Political Corruption, Scandals and Dirty Politics*, political corruption in American predates the origins of the republic. According to Long, corruption during the British administration of the colonies was "routine and not necessarily illegal. That set the stage for an underlying culture of corruption and patronage that ensued after the revolution." I couldn't agree more.

Although political corruption has always been a problem in America it has become so commonplace that many new acts of corruption are not even reported by the media. I would like to say that my experience working as a prosecutor in the Department of Justice made me believe political corruption was under control in the United States. In the far distant past, I did believe this. Today, however, the sheer volume of corruption cases is overwhelming the teams of prosecutors assigned to them. I have to admit that the general view of politicians as distrustful and corrupt is, sadly, a very accurate one.

A.1.2 Control

Prompt:

In the next section you will see an article written in *The New York Times* about Chefs Anthony Bourdain and Eric Ripert. Please pay attention to the article as you will later be asked questions about the content.

Boisson Buddies: Anthony Bourdain and Eric Ripert

On paper they couldn't be more different–one is a refined French chef with four–stars from *The New York Times*, three from the Michelin Guide, and a number of awards from the James Beard Foundation including Top Chef in New York City and Outstanding Chef in the United States; the other is a New York City-born, five-time Emmy-nominated, world-traveling culinary renegade whose first novel, *Bone in the Throat*, is being adapted for the big screen.

Yet Le Bernardin's Eric Ripert and best-selling author and television host Anthony Bourdain go together like moules and frites. Whereas Ripert, his wife, Sandra, and their son, Adrien, have been summering in the Hampton's for more than a decade, this season is the first for Bourdain and his wife, Ottavia, and their daughter, Ariane. But it is a wonderful respite for Bourdain, who spent 260 days last year traveling for his two Travel Channel series, *No Reservations* and *The Layover* and will launch his graphic novel *Get Jiro!* in the fall.

On Eric McDowell shirt, Billy Reid ($175). 54 Bond St., NYC, 212-598-9355. Lightweight slim-fit denim, Lacoste ($150). Americana Manhasset, 2060 Northern Blvd., 516-365-1933. White gold submariner watch, Rolex ($36,850). London Jewlers, 2 Main St., East Hampton, 329-3939.com. Bracelets, Ripert's own. On Anthony: Pinpoint Bengal stripe shirt, Simon Spurr ($225). Singer22, 11 Old Westbury Road, East Hills, 877-474-6722. Sunglasses, khaki's and watch, Bourdain's own.

It was in the least likeliest of ways that these two toques came together. When Bourdain turned the heat up on the culinary industry in his 2001 breakout book *Kitchen Confidential: Adventures in the Culinary Underbelly*, he had many complementary passages about Ripert and Le Bernardin.

"Seventy-five percent of the industry was saying, 'it's scandalous' and 'this guy is a disgrace.' Then part of the industry was saying, 'he's genius,"' remembers Ripert. "I called him and said, 'I read your book, and I would love to know you. Would you come for lunch?' That was the first time I met Anthony, and we have been friends ever since."

Popover shirt, Billy Reid ($185). 54 Bond St., NYC, 212-598-9355. On Anthony: Black T-shirt, stylist's own. Jeans and watch, Bourdain's own.

"When Eric called, the book was doing really well, but I was still working every day at [Brasserie] Les Halles, convinced, quite certain that I should keep my day job and that there was no way that I would be able to support myself or count on writing as an income stream of any kind" says Bourdaindain. "I was absolutely floored that a chef who I respected that much from a restaurant that I never could have been able to afford would call me up and invite me to lunch. When I had the opportunity to get really good at my craft, I chose not to and went the other way. I'm sure in many ways it has been trying to maintain and protect the reputation of an establishment like Le Bernardin and have a friend like me who is likely to put his foot in his mouth every five minutes. It speaks well of Eric's character."

Even stranger, a filmmaker shot me as I'm leaving the lunch, and I'm devastated, just standing there practically in tears. I'd had this amazing meal, and I really saw the road not taken. I had made some very basic decisions about my career, either knowingly or in a calculated way. When I had the opportunity to get really good at my craft, I chose not to and went the other way.

A.1.3 Honest

Prompt:

In the next section you will see an opinion piece about political corruption written in The New York Times by Charles Delauney, an adjunct professor of law at the University of Chicago and former Chief Prosecutor in the United States Department of Justice's Public Integrity Section. Please pay attention to the article as you will later be asked questions about the content.

It Only Seems That Political Corruption is Rampant

Much has been written about political trust lately. It has become very fashionable to call politicians distrustful. Americans, of course, have often been skeptical about government and politicians, but over the last decades, this skepticism has hardened. The recent increase in media coverage of political corruption seems to suggest it is rampant; however, this rise in hardened skepticism is unwarranted. Although public perceptions that politicians are corrupt have skyrocketed, political corruption has in fact decreased over time.

I led my first prosecution of political corruption in 1992 and things have certainly improved since then. Today, political corruption is at an all-time low. According to a recent study by Larry J. Sabato, director of the University of Virginia's Center for Politics, political corruption was much more common throughout the 19th and 20th centuries than today. The idea that aspiring politicians are driven to a career in politics solely for their own personal gain is nonsense. Most politicians work significantly longer hours and get paid considerably less than they would in the private sector. And they face intense and, increasingly, vindictive media scrutiny.

Representative William J. Jefferson was convicted of bribery, racketeering and money laundering in 2009.

It is true that there are occasional corruption scandals involving politicians, but the perception of widespread malfeasance is primarily driven by media scrutiny. In contrast to countries like Italy where political scandals are the norm, we pay them so much attention in the United States precisely because they are so unusual. One occasional rotten apple, or even a handful over the past 20 years, does not spoil the barrel.

Political corruption is simply a rare event in America and although media coverage is easy to find the facts tell a different story. I worked for 25 years as a prosecutor in the Justice Department's public integrity section, which prosecutes official corruption at all levels of government. Although resources devoted to prosecuting corrupt officials have steadily increased since the 1980s, convictions of politicians dropped nearly a quarter from 1989 to 2011. This reflects the fact that only a very few politicians engage in corruption.

Representative John W. Jenrette Jr, right, was convicted in a high-profile scandal

The relatively small percentage of corrupt politicians in the United States is well known by scholars. A 2007 book called *The Almanac of Political Corruption, Scandals and Dirty Politics*, concluded that less than 1 percent of the nearly 12,000 people who had served in Congress had been expelled, indicted or tried for crimes. According to the book's author, Kim Long: "There's a large majority of voters who believe it's just endemic ... There's no evidence that indicates it's the case – zero." I couldn't agree more.

Political corruption has always existed and the only thing rampant is the public misconception about political corruption. The view of corruption as commonplace is not supported by the evidence. In my experience as a prosecutor the facts best represent the reality that political corruption in America is a very rare event. Although it is not reflected in public discussion and media coverage, using words like truthful, honest, sincere, loyal, and genuine to describe most politicians is not as ridiculous as it may sound.

A.2 Experiment 3

Experiment 3 differed from Experiments 1-2 in the following respects:

1. The Op-Eds were presented as excerpts from the print version of the *New York Times*, rather than the online version of the paper. These are presented in full in Figure S19 (*Honest*), Figure S17 (*Corrupt*), and Figure S17 (*Control*).
2. The Op-Eds were supplemented with data visualizations, based on real publicly available data. In the case of the *Honest* and *Corrupt* treatments, these visualizations supported the Op-Ed writer's argument. In the case of the *Control* treatment about recycling, the visualization was domain specific but did not directly refute or undermine the author's argument against the futility of recycling. These are presented in full in Figure S20 (*Honest*), Figure S16 (*Corrupt*), and Figure S18 (*Control*).
3. As in the previous politics experiments, all subjects were provided with an explanatory prompt before receiving treatment. In this experiment, however, subjects were required to stay "in treatment" for at least 60 seconds before proceeding to the rest of the survey:

   > "In the next section, you will see an opinion article published in a major U.S. newspaper. Please read the article carefully. We will ask you some questions about the topic after reading. You may zoom in on the article with your browser. To give you time to read the article, the button to continue to the next question will not appear until after 60 seconds."

4. To encourage and assess engagement with article content, subjects were asked the following questions after reading the randomly assigned article.
   - "What do you think the author's purpose was for writing the piece you just read?" [open-ended]
   - "On a scale of 0 to 100 where 0 is "completely unconvincing" and 100 is "completely convincing", how persuasive did you find the article you just read?"

- "On a scale of 0 to 100 where 0 is "very difficult to understand" and 100 is "very easy to understand", how did you find the author's writing style?"
- "Generally speaking, how interested were you in the topic before reading the article?"
  (a) Very interested (5)
  (b) Somewhat interested (4)
  (c) Neither interested nor uninterested (3)
  (d) Somewhat uninterested (2)
  (e) Very uninterested (1)
- "How likely are you to conduct your own research to find out more about the topic?"
  (a) Very likely (5)
  (b) Somewhat likely (4)
  (c) Neither likely nor unlikely (3)
  (d) Somewhat unlikely (2)
  (e) Very unlikely (1)

  The results are presented in Table S40. Although there were no meaningful differences in perceived accessibility of the Op-Ed writing style across conditions, respondents rated the *Corrupt* Op-Ed as substantially more convincing – by about 20 points on a 100 point scale – than either the *Control* or *Honest* Op-Ed.

5. After subjects responded to these questions, they were provided with the accompanying data visualization and asked two additional questions. Subjects in *Honest* and *Corrupt* were asked:
   (a) "The plot below shows the number of public officials charged with corruption in the United States from 1996 to 2013. How has the number of public officials

charged with corruption changed over time?"

i. Strongly increased (5)
ii. Somewhat increased (4)
iii. Neither increased nor decreased (3)
iv. Somewhat decreased (2)
v. Strongly decreased (1)

(b) "In your opinion, does this plot provide evidence that supports or undermines the author's argument?"

i. Strongly supports (5)
ii. Somewhat supports (4)
iii. Neither supports not undermines (3)
iv. Somewhat undermines (2)
v. Strongly undermines (1)

Subjects in *Control* were asked the same two questions, but the language in the first was changed to match the content: "The plot below shows the total amount of unrecycled municipal waste in the United States from 1960 to 2010. How has the amount of unrecycled municipal waste changed over the time period?" The results presented graphically in Figures S13 and S14. Point estimates and standard errors are presented in Table S41. Trends were clearly interpreted in the correct direction. Respondents believed the data visualization in the *Corrupt* condition was more supportive of the writer's argument than the data visualization in the *Honest* condition.

6. As in the previous experiments, all subjects were debriefed about the deception involved at the end of the experiment. The author had moved to a new university at the time this experiment was conducted. The Institutional Review Board at this university did not require that subjects be given the option of removing their data upon learning of

the deception. Figure S11 and Figure S12 show the debrief information that subjects assigned to the *Honest* or *Corrupt* conditions received. Subjects assigned to the *Control* condition did not receive a debrief since there was no deception–the content was copied directly from a *New York Times* Op-Ed on unrecycled municipal waste, and credited as such.

Figure S11: Debrief in Experiment 3

Thank you for your participation in this study! The goal of this study was to understand shifts in public opinion based on certain newspaper articles and viewpoints expressed by writers. For this reason, the article you read was not in fact published in a major U.S. newspaper. Instead, it was created for the purposes of this study, but also loosely based on past and current events.

We also showed you a graphic depicting changes in charges of corruption against public officials over time in the United States. Each year, the FBI delivers a report to Congress on the activities and operations of the Public Integrity Section, a division that prosecutes public officials for corruption. The graphic you saw was created using data published in a recent FBI report. It's important to note that **although charges against federal officials have decreased over the period 1996-2013, charges against state and local officials have increased.**

The graphic below shows the data from the FBI report with different trend lines for each level of government. Averaging over all levels of government, charges of corruption are essentially constant (no change) over the period 1996-2013. You can learn more about the FBI office that prosecutes public officials for corruption at https://www.justice.gov/criminal/pin.

Figure S12: Trends in DOJ Prosecutions of Public Officials by Level of Government

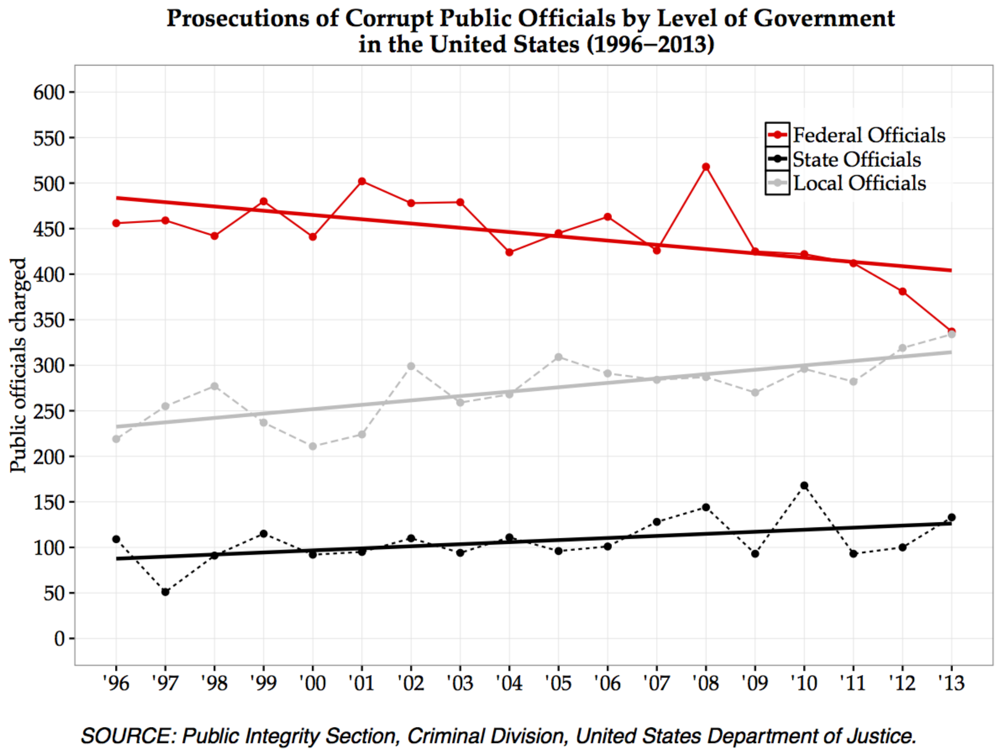


Table S40: Group means for Questions about Op-Eds

	Corrupt	Control	Honest
Accessible	78.89 (0.94)	78.22 (1.05)	80.42 (1.01)
Convincing	75.98 (0.77)	54.26 (1.14)	52.44 (1.17)
Own Research	3.27 (0.05)	3.40 (0.05)	3.12 (0.05)
Prior Interest	3.79 (0.04)	3.75 (0.05)	3.60 (0.04)

Table S41: Group means for Questions about Data Visualizations

	Corrupt	Control	Honest
Evidence supports	4.56 (0.03)	3.34 (0.04)	4.11 (0.04)
Trend increased	4.69 (0.02)	4.76 (0.03)	1.54 (0.03)

Figure S13: Interpretation of trends by treatment arm

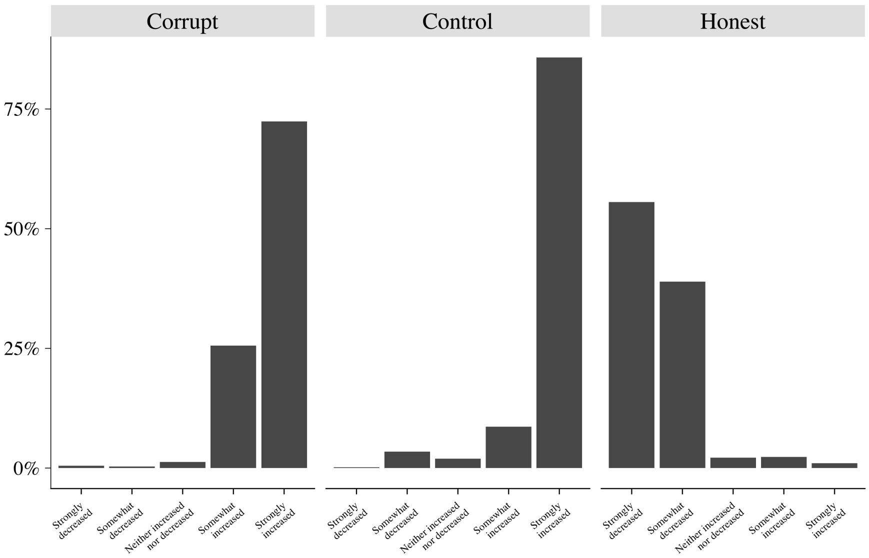
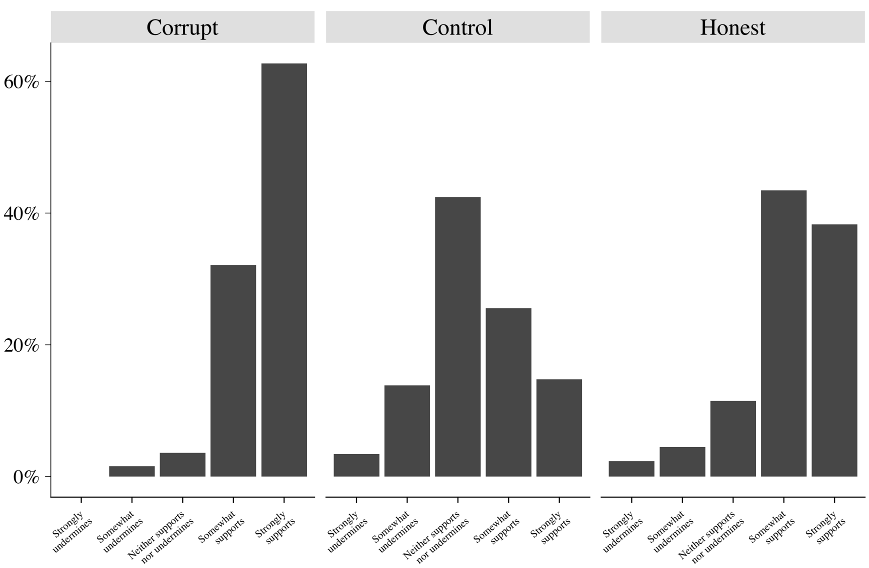


Figure S14: Interpretation of support by treatment arm



A.2.1 Corrupt

Figure S15: Corrupt Op-Ed in Experiment 3

Political Corruption Is More Widespread Than It Seems

By PATRICK J. FITZGERALD

MUCH has been said lately about political corruption and it has become very fashionable to call politicians corrupt and untrustworthy. Americans, of course, have always been skeptical of government and politicians. Over the last decade, however, this skepticism has hardened.

This rise in hardened skepticism is warranted. Media coverage of political corruption has downplayed the severity of the problem. Although political corruption is more widespread than ever, the public's perception is that corruption in America has significantly decreased over time.

I led my first prosecution of political corruption in 1992 when an F.B.I. sting involving bribery and horseracing legislation netted convictions of nearly 11 percent of the Kentucky Legislature. Things have gotten even worse since then. Today, political corruption is at an all-time high.

According to a recent study by Larry J. Sabato, director of the University of Virginia's Center for Politics, political corruption was much less common throughout the 19th and 20th centuries than today. The idea that aspiring politicians are driven to a career in politics solely to serve the public is nonsense.

Most politicians spend over half their workday asking for campaign donations rather than performing the jobs they were elected to do. They earn a six-figure salary while working for the people "part-time".

It is true that there are occasional news stories of corruption scandals involving politicians, but the true level of widespread malfeasance is hidden by a lack of media scrutiny. According to the anti-corruption organization Transparency International, the United States has one of the highest levels of corruption in the developed world. In contrast to countries like Denmark or New Zealand, where political corruption is rare, we pay so little attention to corruption in the United States precisely because it is so common.

I worked for over 20 years in the Justice Department, and investigated official corruption at all levels of government. Although resources devoted to prosecuting corrupt officials have steadily decreased since the 1980s, convictions of politicians have increased more than a quarter since 1989. This reflects the fact that prosecutors have had to do more with less.

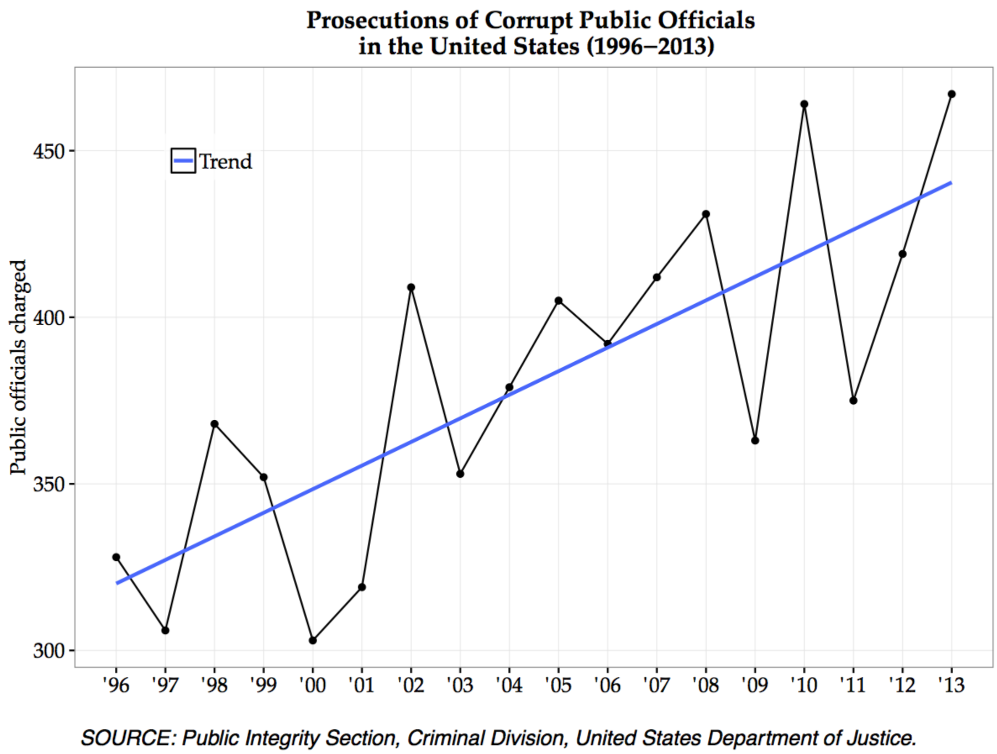
The relatively large percentage of corrupt politicians in the United States is well known by scholars. A 2007 book called *The Almanac of Political Corruption, Scandals and Dirty Politics*, concluded that political corruption in America predates the origins of the republic. According to the book's author, Kim Long, corruption during the British administration of the colonies was routine: "That set the stage for an underlying culture of corruption and patronage that ensued after the revolution." I couldn't agree more.

We have always had a culture of corruption in this country but it's now worse than ever. In 2011, I prosecuted Rod Blagojevich – the fourth of the last seven Illinois Governors to go to prison. Today, Gov. Terry McAuliffe of Virginia is facing a federal investigation. If charged, he would be the tenth governor or former governor since 2000.

I wish I could say that my experience as a prosecutor made me believe political corruption was under control in the United States. In the far distant past, I did believe this. Today, I believe the view that most politicians are corrupt is sadly, a very accurate one.

Patrick J. Fitzgerald is the former U.S. Attorney for the Northern District of Illinois and partner at the Chicago office of Skadden, Arps, Slate, Meagher & Flom.

Figure S16: Data Visualization of increasing Corruption in Experiment 3



A.2.2 Control

Figure S17: Control Op-Ed in Experiment 3

Recycling Has Failed in Economic and Environmental Terms

By JOHN TIERNEY

IF you live in the United States, you probably do some form of recycling. It's likely that you separate paper from plastic and glass and metal. You rinse the bottles and cans, and you might put food scraps in a container destined for a composting facility.

As you sort everything into the right bins, you probably assume that recycling is helping your community and protecting the environment. But is it? Are you in fact wasting your time?

In 1996, I wrote an article for The New York Times Magazine arguing that the recycling process as we carried it out was wasteful. I presented plenty of evidence that recycling was costly and ineffectual, but its defenders said that it was unfair to rush to judgment. Noting that the modern recycling movement had really just begun just a few years earlier, they predicted it would flourish as the industry matured and the public learned how to recycle properly.

As cities move beyond recycling paper and metals, and into glass, food scraps and assorted plastics, the costs rise sharply while the environmental benefits decline and sometimes vanish. "If you believe recycling is good for the planet and that we need to do more of it, then there's a crisis to confront," says David P. Steiner, the chief executive officer of Waste Management, the largest recycler of household trash in the United States. "Trying to turn garbage into gold costs a lot more than expected. We need to ask ourselves: What is the goal here?"

Recycling has been relentlessly promoted as a goal in and of itself: an unalloyed public good and private virtue that is indoctrinated in students from kindergarten through college. As a result, otherwise well-informed and educated people have no idea of the relative costs and benefits.

They probably don't know, for instance, that to reduce carbon emissions, you'll accomplish a lot more by sorting paper and aluminum cans than by worrying about yogurt containers and half-eaten slices of pizza. Most people also assume that recycling plastic bottles must be doing lots for the planet.

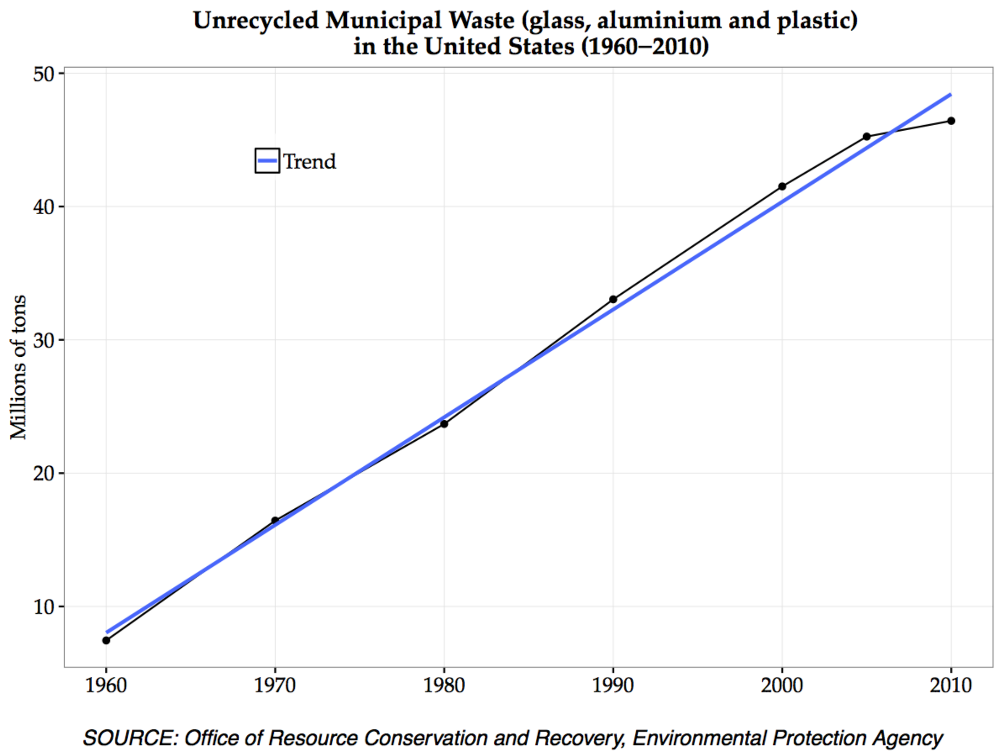
But how much difference does it make? Here's some perspective: To offset the greenhouse impact of one passenger's round-trip flight between New York and London, you'd have to recycle roughly 40,000 plastic bottles, assuming you fly coach. If you sit in business- or first-class, where each passenger takes up more space, it could be more like 100,000.

The environmental benefits of recycling come chiefly from reducing the need to manufacture new products — less mining, drilling and logging. But that's not so appealing to the workers in those industries and to the communities that have accepted the environmental trade-offs that come with those jobs.

As a business, recycling is on the wrong side of two long-term global economic trends. For centuries, the real cost of labor has been increasing while the real cost of raw materials has been declining. That's why we can afford to buy so much more stuff than our ancestors could. As a labor-intensive activity, recycling is an increasingly expensive way to produce materials that are less and less valuable.

John Tierney is the writer of the Findings column for The New York Times Science section and co-author of the book "Willpower: Rediscovering the Greatest Human Strength."

Figure S18: Data Visualization of increasing Municipal Waste in Experiment 3



A.2.3 Honest

Figure S19: Honest Op-Ed in Experiment 3

It Only Seems That Political Corruption Is Widespread

By PATRICK J. FITZGERALD

MUCH has been said lately about political corruption and it has become very fashionable to call politicians corrupt and untrustworthy. Americans, of course, have always been skeptical of government and politicians. Over the last decade, however, this skepticism has hardened.

This rise in hardened skepticism is unwarranted. Media coverage of political corruption has exaggerated the severity of the problem. Although the public's perception is that political corruption in America has skyrocketed, it has significantly decreased over time.

I led my first prosecution of political corruption in 1992 when an F.B.I. sting involving bribery and horseracing legislation netted convictions of nearly 11 percent of the Kentucky Legislature. Things have certainly improved since then. Today, political corruption is at an all-time low.

According to a recent study by Larry J. Sabato, director of the University of Virginia's Center for Politics, political corruption was much more common throughout the 19th and 20th centuries than today. The idea that aspiring politicians are driven to a career in politics solely for their own personal gain is nonsense.

Most politicians work significantly longer hours and get paid considerably less than they would in the private sector. They also face intense and increasingly vindictive media scrutiny.

It is true that there are occasional corruption scandals involving politicians, but the perception of widespread malfeasance is primarily driven by media scrutiny. According to the anticorruption organization Transparency International, the United States has one of the lowest levels of corruption in the world. In contrast to countries like China or Nigeria, where political corruption is the norm, we pay so much attention to corruption in the United States precisely because it is so unusual.

I worked for over 20 years in the Justice Department, and investigated official corruption at all levels of government. Although resources devoted to prosecuting corrupt officials have steadily increased since the 1980s, convictions of politicians have dropped nearly a quarter since 1989. This reflects the fact that only a very few politicians engage in corruption.

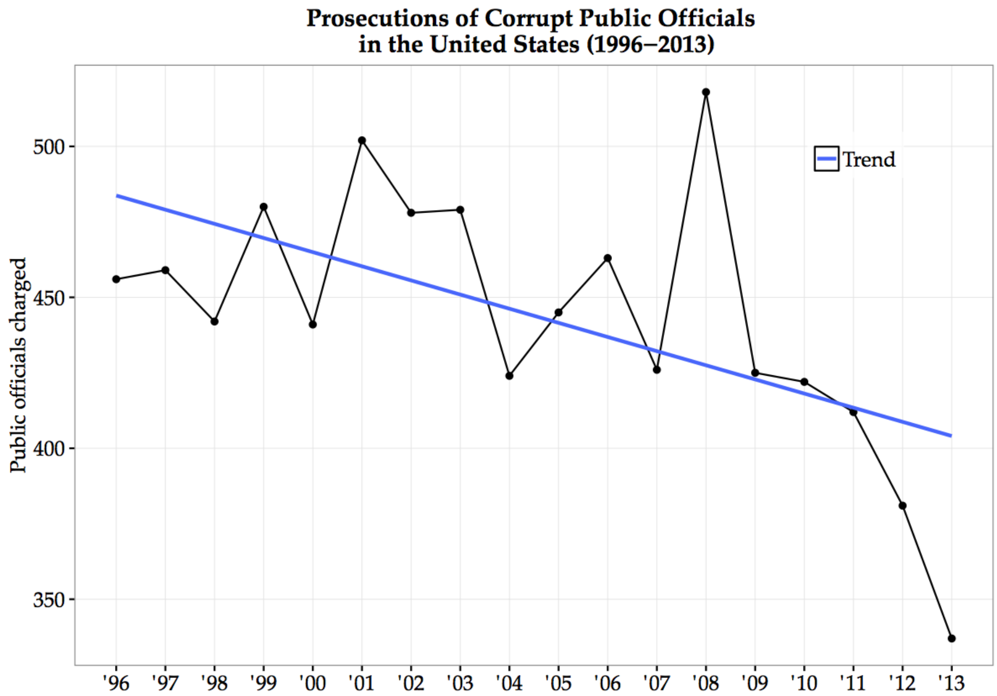
The relatively small percentage of corrupt politicians in the United States is well known by scholars. A 2007 book called *The Almanac of Political Corruption, Scandals and Dirty Politics*, concluded that less than 1 percent of the nearly 12,000 people who had served in Congress had been expelled, indicted or tried for crimes. According to the book's author, Kim Long: "There's a large majority of voters who believe it's just endemic ... There's no evidence that indicates it's the case -- zero." I could not agree more.

One occasional rotten apple, or even a handful over the past 20 years, does not spoil the barrel. Political corruption has always existed at some small level. The only thing rampant is the public misconception about political corruption. The evidence simply does not support the view that corruption is commonplace.

In my experience as a prosecutor the facts best represent the reality that political corruption in America is a very rare event. Although it is not reflected in public discussion and media coverage, using words like truthful, honest, sincere, loyal, and genuine to describe most politicians is not as ridiculous as it may sound.

Patrick J. Fitzgerald is the former U.S. Attorney for the Northern District of Illinois and partner at the Chicago office of Skadden, Arps, Slate, Meagher & Flom.

Figure S20: Data Visualization of decreasing Corruption in Experiment 3



SOURCE: Public Integrity Section, Criminal Division, United States Department of Justice.

A.3 Experiment 4-5

A.3.1 Corrupt

Prompt:

> In the next section you will see an opinion piece about corruption in the National Football League (NFL) written in *The New York Times* by Charles Delauney, an adjunct professor of law at the University of Chicago and former Chief Prosecutor in the United States Department of Justice's Public Integrity Section. Please pay attention to the article as you will later be asked questions about the content.

Corruption in the National Football League is rampant.

Much has been written about trust and the National Football League (NFL) recently and it has become very fashionable to call NFL players distrustful. Prominent scandals have always been part of American sports leagues. Rumors that the White Sox threw the 1919 World Series resulted in a national crisis as owners of Major League Baseball (MLB) teams worried the game lost the public trust. This lack of trust seems to have hit the NFL in recent years and the integrity of NFL officials and players is worse than ever.

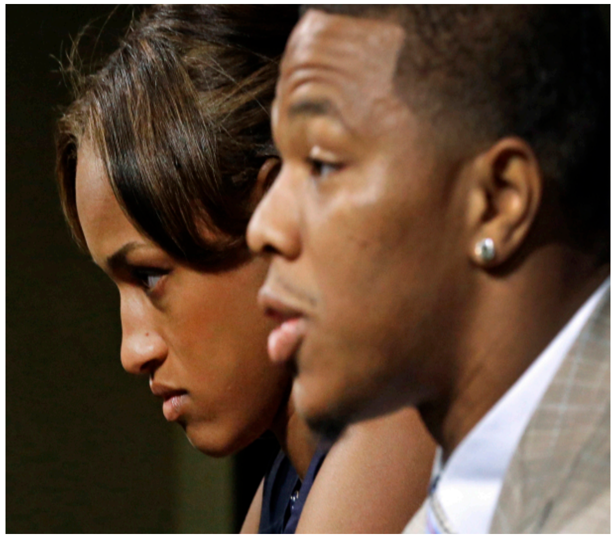
Today we see cases of bad behavior by NFL players and officials in increasing scope and frequency. For example, the NFL recently suspended Baltimore Ravens player Ray Rice after he was indicted for knocking his wife unconscious in a New Jersey casino. Video footage released approximately five months later showed Rice knocking out his wife and dragging her unconscious body from an elevator. Some journalists speculated NFL commissioner Goodell attempted to cover up whether the NFL knew about the video and the National Organization for Women called for his resignation.

NFL Comissioner Roger Goodell faces criticism over the NFL's response to player misconduct.

This case highlights a widespread problem. According to a report by The New York Times, domestic violence in the NFL is extraordinarily high. Based on data from the Bureau of Justice Statistics Report and the USA Today NFL Arrests Database, the arrest rate for domestic violence among NFL players is significantly higher than men of comparable age and income.

Players are also frequently arrested for driving under the influence (DUI), assault and weapons charges yet typically avoid serious punishment. In 1998, Leonard Little of the St. Louis Rams crashed into and killed Susan Gutweiler. He had a blood alcohol level over twice the legal limit and received four years probation and 1000 hours of community service.

Little was suspended for eight games and arrested again in 2004 for DUI. He received two years probation and continued to play for the Rams until 2009. The NFL does not take player misconduct seriously at it never has. The unscrupulous reputation players have earned is well deserved and NFL corruption is at an all time high.



Baltimore Ravens player Ray Rice and his wife Janay Rice at a news conference in May 2014. Rice was indefinitely suspended from the NFL in September 2014.

Incidents of player misconduct are commonplace and fostered by the NFL's disregard for standards of decency and a culture of corruption. As Pulitzer Prize winning sportswriter Thomas Boswell observed: "The NFL is the league where you hold your breath week to week, almost day to day, to find out what crime, what betrayal of trust, what warped values for the young the sport can become identified with next."

I couldn't agree more. The public perception that many NFL players are criminals is correct. Corruption and misconduct has become so commonplace that many new acts are not even reported by the media. In my experience as a prosecutor I would like to say the problem is exaggerated. In the far distant past I did believe this. Today, however, I have to admit that the general view of the NFL and its players as distrustful and corrupt is, sadly, a very accurate one.

A.3.2 Control

Identical to *Control* in Experiments 1-2

A.3.3 Honest

Prompt:

> In the next section you will see an opinion piece about corruption in the National Football League (NFL) written in *The New York Times* by Charles Delauney, an adjunct professor of law at the University of Chicago and former Chief Prosecutor in the United States Department of Justice's Public Integrity Section. Please pay attention to the article as you will later be asked questions about the content.

It only seems that Corruption in the National Football League is rampant.

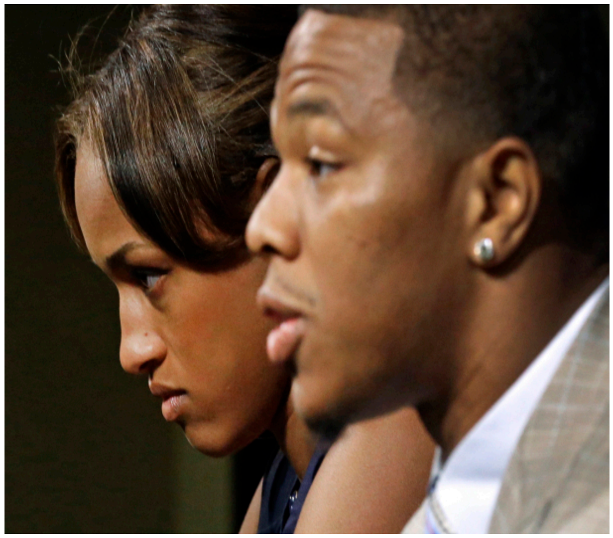
Much has been written about trust and the National Football League (NFL) recently and it has become very fashionable to call NFL players distrustful. Prominent scandals have always been part of American sports leagues. Rumors that the White Sox threw the 1919 World Series resulted in a national crisis as owners of Major League Baseball (MLB) teams worried the game lost the public trust. This lack of trust seems to have hit the NFL in recent years, yet the integrity of NFL officials and players is better than ever.

According to Larry Sabato, director of the NFL Arrests Database at the University of Virginia, poor behavior by players was much more common 50 years ago than today. Despite what media reports claim, most major league players are responsible citizens. Although public perceptions that the NFL and its players are corrupt have skyrocketed, the data suggest the opposite.

NFL Comissioner Roger Goodell faces criticism over the NFL's response to player misconduct.

The endless media stories about the NFL simply reflect the fact that sports players are under constant media scrutiny. There is no evidence to suggest anything unusual about NFL players. On the contrary, a report by *The New York Times* showed arrest rates among players are extremely low compared to national averages. Based on data from the Bureau of Justice Statistics Report and the USA Today NFL Arrests Database, the arrest rate for players is only 13% of the national average for men of comparable age.

The real trend is that players are swiftly punished for their behavior off the field. For example, Viking's Adrian Peterson was recently suspended without pay until April 2015 following allegations of whipping his child with a tree branch. Suspensions of this length prior to conviction are unprecedented. Consider the 1998 case of St. Louis Rams' Leonard Little, who crashed into and killed Susan Gutweiler. He had a blood alcohol level over twice the legal limit and received four years probation and community service. He was only suspended for eight games and played for the Rams until 2009. Punishment has become more severe since then and violent behavior is down as a consequence.



Baltimore Ravens player Ray Rice and his wife Janay Rice at a news conference in May 2014. Rice was indefinitely suspended from the NFL in September 2014.

Player misconduct is increasingly rare in both occurrence and severity. In 1999 Alfred Blumstein, a leading criminologist, compared rates of criminal violence among NFL players to the general population and found that the annual rate of assault and domestic violence by NFL players was less than half the national average. The recent *The New York Times* report shows incidents of player misconduct have decreased over time. Blumstein and colleagues have repeated shown that NFL players are less violent than the general population and that player conduct has not deteriorated over time.

I couldn't agree more. It seems that the only thing rampant is the public perception that NFL players are criminals. In my experience as a prosecutor the statistics cited above represent the reality that bad behavior by NFL players and officials is very rare. Although not reflected in public discussion and media coverage, words like truthful, honest, sincere, loyal, and genuine describe most players and NFL officials.

Bibliography

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Ansolabehere, Stephen, Jonathan Rodden and James M Snyder. 2008. "The strength of issues." *American Political Science Review* 102(2):215–232.

Aronow, Peter M., Jonathon Baron and Lauren Pinson. 2019. "A Note on Dropping Experimental Subjects who Fail a Manipulation Check." *Political Analysis* p. 1–18.

Athey, Susan, Julie Tibshirani and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47(2):1148–1178.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.

Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20:351–368.

Bischoff, Ivo, Björn Frank et al. 2011. "Good news for experimenters: Subjects are hard to influence by instructors' cues." *Economics Bulletin* 31(4):3221–3225.

Chandler, Jesse, Gabriele Paolacci, Eyal Peer, Pam Mueller and Kate A Ratliff. 2015. "Using nonnaive participants can reduce effect sizes." *Psychological science* 26(7):1131–1139.

Ciolli, Joe. 2017. "Sinking NFL viewership is threatening to crush ad sales." *Business Insider* .
URL: *http://www.businessinsider.com/sinking-nfl-viewership-is-threatening-to-crush-ad-sales-2017-12*

Coppock, Alexander. 2018. *ri2: Randomization Inference for Randomized Experiments*. R package version 0.1.1.
URL: *https://CRAN.R-project.org/package=ri2*

Coppock, Alexander, Thomas J Leeper and Kevin J Mullinix. 2018. "Generalizability of heterogeneous treatment effect estimates across samples." *Proceedings of the National Academy of Sciences* 115(49):12441–12446.

De Quidt, Jonathan, Johannes Haushofer and Christopher Roth. 2018. "Measuring and bounding experimenter demand." *American Economic Review* 108(11):3266–3302.

Doherty, Carroll, Jocelyn Kiley, Alec Tyson and Bridget Jameson. 2015. "Beyond distrust: How Americans view their government." *Pew Research Center* .
URL: *http://www.people-press.org/files/2015/11/11-23-2015-Governance-release.pdf*

Dunn, Thomas J., Thom Baguley and Vivienne Brunsden. 2013. "From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation." *British Journal of Psychology* .
URL: *http://dx.doi.org/10.1111/bjop.12046*

Faulkner, Nicholas, Aaron Martin and Kyle Peyton. 2015. "Priming political trust." *Australian Journal of Political Science* 50(1):164–173.

Gerber, Alan S and Donald P Green. 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gershtenson, Joseph and Dennis L Plane. 2007. "Trust in government." *American National Election Studies pilot report* .

Glass, Gene V. 1976. "Primary, secondary, and meta-analysis of research." *Educational researcher* 5(10):3–8.

Hartman, Erin and F Daniel Hidalgo. 2018. "An equivalence approach to balance and placebo tests." *American Journal of Political Science* 62(4):1000–1013.

Hauser, David J and Norbert Schwarz. 2016. "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants." *Behavior research methods* 48(1):400–407.

Hetherington, Marc J. 2005. *Why Trust Matters*. Princeton: Princeton University Press.

Interactive, Harris. 2015. "If you had to choose, which one of these sports would you say is your favorite?" *Statista – The Statistics Portal.* .
URL: *https://www.statista.com/statistics/198409/us-adults-favorite-sports/*

Kam, Cindy D and Marc J Trussler. 2017. "At the nexus of observational and experimental research: Theory, specification, and analysis of experiments with heterogeneous treatment effects." *Political Behavior* 39(4):789–815.

Kelley, Ken and Keke Lai. 2012. *MBESS: MBESS*. R package version 3.3.3.
URL: *http://CRAN.R-project.org/package=MBESS*

Kreuter, Frauke, Stanley Presser and Roger Tourangeau. 2008. "Social desirability bias in CATI, IVR, and Web surveys the effects of mode and question sensitivity." *Public Opinion Quarterly* 72(5):847–865.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez and Stefanie Stantcheva. 2015. "How elastic are preferences for redistribution?" *The American Economic Review* 105(4):1478–1508.

Lakens, Daniël. 2017. "Equivalence tests." *Social Psychological and Personality Science* 8(4):355–362.

Lakens, Daniël, Anne M Scheel and Peder M Isager. 2018. "Equivalence testing for psychological research: A tutorial." *Advances in Methods and Practices in Psychological Science* 1(2):259–269.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Orne, Martin T. 1962. "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications." *American psychologist* 17(11):776.

Paolacci, Gabriele, Jesse Chandler, Panagiotis G Ipeirotis et al. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5):411–419.

Peer, Eyal, Joachim Vosgerau and Alessandro Acquisti. 2014. "Reputation as a sufficient condition for data quality on Amazon Mechanical Turk." *Behavior research methods* 46(4):1023–1031.

Rainey, Carlisle. 2014. "Arguing for a negligible effect." *American Journal of Political Science* 58(4):1083–1091.

Rudolph, Thomas J. 2017. Political Trust as a Heuristic. In *The Handbook of Political Trust*, ed. Sonja Zmerli and Tom var der Meer. Cheltenham: Edward Algar Publishing pp. 197–211.

Stock, James H, Jonathan H Wright and Motohiro Yogo. 2002. "A survey of weak instruments and weak identification in generalized method of moments." *Journal of Business & Economic Statistics* 20(4):518–529.

Tella, Rafael Di, Juan Dubra and Alejandro Luis Lagomarsino. 2016. Meet the oligarchs: Business legitimacy, state capacity and taxation. Technical report National Bureau of Economic Research.

Wellek, Stefan. 2010. *Testing statistical hypotheses of equivalence and noninferiority*. Chapman and Hall/CRC.

White, Ariel, Anton Strezhnev, Christopher Lucas, Dominika Kruszewska and Connor Huff. 2018. "Investigator characteristics and respondent behavior in online surveys." *Journal of Experimental Political Science* 5(1):56–67.

Wines, Michael. 2014. "It Only Seems That Political Corruption Is Rampant." *The New York Times* .
URL: *https://www.nytimes.com/2014/01/26/us/politics/it-only-seems-that-political-corruption-is-rampant.html*