Corruption information and vote share: A meta-analysis and lessons for survey experiments*

Trevor Incerti†

Yale University

May 19, 2019

Abstract

Do voters hold politicians accountable for corruption? Field experiments that provide voters with information about the corrupt acts of politicians then monitor vote choice have become standard. Similarly, vote choice survey experiments commonly inform respondents of the corrupt acts of hypothetical candidates. What have we learned from these experiments? Meta-analysis demonstrates that the aggregate treatment effect of providing information about corruption on vote share in field experiments is approximately zero. By contrast, corrupt candidates are punished by respondents by approximately 33-35 percentage points across survey experiments. This suggests that while vote-choice survey experiments may provide information on the directionality of informational treatments in idealized hypothetical scenarios, the point estimates they provide may not be representative of real-world voting behavior. I explore publication bias, social desirability bias, and the nature of the experimental designs as potential explanations for this discrepancy. Finally, I offer suggestions for analyzing vote-choice conjoint experiments when researchers have strong theories about the conditions that shape voter decision-making.

First draft: March 2019. This draft: May 2019

[PRELIMINARY DRAFT: ADDITIONAL STUDIES TO BE ADDED AND POINT ESTIMATES REFINED. PLEASE DO NOT CITE OR CIRCULATE WITHOUT AUTHOR'S PERMISSION.]

*

†trevor.incerti@yale.edu

1 Introduction

Competitive elections create a system whereby voters can hold policy makers accountable for their actions. This mechanism should make politicians hesitant to engage in malfeasance such as blatant acts of corruption. Increases in public information regarding corruption should therefore decrease levels of corruption in government, as voters armed with information should expel corrupt politicians (Gray & Kaufman 1998; Kolstad & Wiig 2009; Rose-Ackerman & Palifka 2016). However, this theoretical prediction is undermined by the observation that well-informed voters continue to vote corrupt politicians into office in many democratic states. Political scientists and economists have therefore turned to experimental methods to test the causal effect of learning about politician corruption on vote choice.

Numerous experiments have examined whether providing voters with information about the corrupt acts of politicians decreases their re-election rates. These papers often suggest that there is little consensus on how voters respond to information about corrupt politicians (Arias, Larreguy, Marshall, & Querubin 2018; Botero, Cornejo, Gamboa, Pavao, & Nickerson 2015; Buntaine, Jablonski, Nielson, & Pickering 2018; De Vries & Solaz 2017; Klašnja, Lupu, & Tucker 2017; Solaz, De Vries, & de Geus 2018). Others indicate that experiments have provided us with evidence that voters strongly punish individual politicians involved in malfeasance (Chong, De La O, Karlan, & Wantchekon 2014; Weitz-Shapiro & Winters 2017; Winters & Weitz-Shapiro 2015, 2016).

By contrast, this meta-analysis suggests that: (1) In aggregate, the effect of providing information about incumbent corruption on incumbent vote share in field experiments is approximately zero, and (2) corrupt candidates are punished by respondents by approximately 33-35 percentage points across survey experiments. I also examine mechanisms that may give rise to this discrepancy. I do not find systematic evidence of publication bias. Social desirability bias may lead survey experiments to capture anti-corruption norms rather than realistic voter behavior. Field and survey experiments also may be measuring different causal estimands, as field experiments possess the option of voting for alternative candidates or abstaining. Conjoint experiments attempt to alleviate this issue, but are often analyzed in ways that may fail to illuminate the most substantively important comparisons.

2 Corruption information and electoral accountability

Experimental support for the hypothesis that providing voters with information about politicians' corrupt acts decreases their re-election rates is mixed. Field experiments have provided some causal evidence that informing voters of candidate corruption has negative (but generally small) effects on candidate vote-share. This information has been provided by: randomized financial audits (Ferraz & Finan 2008), fliers revealing corrupt actions of politicians (Chong et al. 2014; De Figueiredo, Hidalgo, & Kasahara 2011), and SMS messages (Buntaine et al. 2018). However, near-zero and null findings are also prevalent, and the negative and significant effects reported above sometimes only manifest in particular subgroups. Banerjee, Green, Green, and Pande (2010) primed voters in rural India not to vote for corrupt candidates, and Banerjee, Kumar, Pande, and Su (2011) provided information on politicians' spending discrepancies, with both studies finding near-zero and null effects on vote share. Boas, Hidalgo, and Melo (2018) similarly find zero and null effects from distributing fliers in Brazil. Finally, Arias et al. (2018); Arias, Larreguy, Marshall, and Querubin (2019) find that providing Mexican voters with information (fliers) about mayoral corruption actually *increased* incumbent party vote share by 3%.[1]

By contrast, online survey experiments consistently show large negative effects from information treatments on hypothetical vote share. These experiments often manipulate moderating factors other than information provision (e.g. quality of information, source of information, whether the candidate is a co-partisan or co-ethnic, whether corruption brings economic benefits, etc.), but even so systematically show negative treatment effects (Anduiza, Gallego, & Muñoz 2013; Avenburg 2016; Banerjee, Green, McManus, & Pande

[1]The authors theorize that this average effect stems from levels of reported malfeasance actually being lower than voters no-information expectations of corruption.

2014; Boas et al. 2018; Breitenstein 2019; Eggers, Vivyan, & Wagner 2018; Franchino & Zucchini 2015; Klašnja et al. 2017; Klašnja & Tucker 2013; Mares & Visconti 2019; Vera Rojas 2017; Weitz-Shapiro & Winters 2017; Winters & Weitz-Shapiro 2013, 2015, 2016, 2018). These experiments have historically taken the form of single treatment arm or multiple arm factorial vignettes, but more recently have tended toward conjoint experiments (Breitenstein 2019; Chauchard, Klasnja, & Harish 2017; Franchino & Zucchini 2015; Klašnja et al. 2017; Mares & Visconti 2019).

Boas et al. (2018) find differential results that they obtain from a field and survey experiment—zero and null in field, large and negative in survey. They argue that this may reflect that norms against malfeasance in Brazil do not translate into action in real life. Boas et al. (2018) point to features specific to Brazil in their explanation of this discrepancy, namely lower salience of corruption to Brazilian voters in municipal elections and the strong effects of dynastic politics. However, meta-analysis demonstrates that this is not only the case for Boas et al. (2018)'s experiments in Brazil, but extends across a systematic review of all studies conducted to date. This suggests that the discrepancy between field and survey experimental findings is driven by methodological differences, rather than context.

Lab experiments that reveal corrupt actions to fellow players appear to be similar to survey experiments, and also show large negative treatment effects (see Figure A.1). While there are not enough lab experiments examining whether the provision of corruption information impacts vote choice to conduct a formal meta-analysis (Arvate & Mittlaender 2017; Azfar & Nelson 2007; Rundquist, Strom, & Peters 1977; Solaz et al. 2018), this discrepancy is worth noting as previous examinations of lab-field correspondence have found evidence of general replicability (Camerer 2011; Coppock & Green 2015).

3 Research Design and Methods

3.1 Selection criteria

I followed standard practices to locate the experiments included in the meta-analysis. This included following citation chains and searches of data bases using the terms ("corruption experiment," "corruption field experiment," "corruption survey experiment," "corruption factorial", "corruption candidate choice", "corruption conjoint", "corruption, vote, experiment", and "corruption vignette"). Papers from any discipline are eligible for inclusion, but in practice stem only from economics and political science. Both published articles and working papers are included so as to ensure the meta-analysis is not biased towards published results. In total, I located 10 field experiments from 8 papers, and 18 survey experiments from 15 papers.

Field experiments are included if researchers randomly assigned information regarding incumbent corruption to voters, then measured corresponding voting outcomes. This therefore excludes experiments that randomly assign corruption information, but use favorability ratings or other metrics rather than actual vote share as their dependent variable (Green, Zelizer, Kirby, et al. 2018). I include one natural experiment, Ferraz and Finan (2008), as random assignment was conducted by the Brazilian government. Effects reported in the meta-analysis come from information treatments on the entire sample of study only, not subgroup or interactive effects that reveal the largest treatment effects.

For survey experiments, studies must test a no-information control group versus a corruption information treatment group and measure vote choice for a hypothetical candidate. This necessarily excludes studies that compare one type of information provision (e.g. source) to another and the control group is one type of information rather than no information, or where the politician is always known to be corrupt (Anduiza et al. 2013; Botero et al. 2015; Konstantinidis & Xezonakis 2013; Muñoz, Anduiza, & Gallego 2012; Rundquist et al. 1977; Weschle 2016). In many cases, studies have multiple corruption treatments (e.g. high quality information vs. low quality information, co-partisan vs. opposition party, etc.). In these cases, I replicate the studies and code corruption as a binary treatment (0 = clean, 1 = corrupt) where *all* treatment arms that provide corruption information are combined into a single treatment. Studies that use non-binary vote choices are rescaled into a binary vote choice.[2] In some cases (5 total), point estimates, standard errors and/or confidence intervals are not explicitly reported, and standard errors are estimated by digitally measuring coefficient plots.[3]

3.2 Included studies

A list of all papers - disaggregated by field and survey experiments - that meet the criteria outlined above are provided in Table 1 and Table 2. A list of lab experiments (4 total) can also be found in and Table A.1, although these studies are not included in the meta-analysis. A list of excluded studies with justification for their exclusion can be found in Table A.2.

Table 1: Field experiments

Study	Country	Treatment
Arias et al. (2018)	Mexico	Fliers
Banerjee et al. (2010)	India	Newspaper
Banerjee et al. (2011)	India	Canvas/Newspaper
Boas et al. (2018)	Brazil	Fliers
Buntaine et al. (2018)	Ghana	SMS
Chong et al. (2014)	Mexico	Fliers
De Figueiredo et al. (2011)	Brazil	Fliers
Ferraz and Finan (2008)	Brazil	Audits

[2] For example, a 1-4 scale is recoded so that 1 or 2 is equal to no vote, and 3 or 4 is equal to a vote.

[3] I recognize that this introduces non-statistical measurement error into the meta-analysis. However, it is not possible for these errors to be large enough to effect the substantive conclusions of the analysis.

Table 2: **Survey experiments**

Study	Country	Type of survey
Avenburg (2016)	Brazil	Vignette
Banerjee et al. (2014)	India	Vignette
Breitenstein (2019)	Spain	Conjoint
Boas et al. (2018)	Brazil	Vignette
Chauchard et al. (2017)	India	Conjoint
Eggers et al. (2018)	UK	Conjoint
Franchino and Zucchini (2015)	Italy	Conjoint
Klašnja and Tucker (2013)	Sweden	Vignette
Klašnja and Tucker (2013)	Moldova	Vignette
Klašnja et al. (2017)	Argentina	Conjoint
Klašnja et al. (2017)	Chile	Conjoint
Klašnja et al. (2017)	Uruguay	Conjoint
Mares and Visconti (2019)	Romania	Conjoint
Vera Rojas (2017)	Peru	Vignette
Winters and Weitz-Shapiro (2013)	Brazil	Vignette
Winters and Weitz-Shapiro (2015)	Brazil	Vignette
Winters and Weitz-Shapiro (2016)	Brazil	Vignette
Weitz-Shapiro and Winters (2017)	Brazil	Vignette
Winters and Weitz-Shapiro (2018)	Argentina	Vignette

3.3 Additional selection comments

Additional justification for the inclusion or exclusion of certain studies, as well as coding and/or replication choices may be warranted in some cases. The field experiment conducted by Banerjee et al. (2010) is included. However, the authors treated voters with a campaign not to vote for corrupt candidates, but did not provide voters with information on which candidates were corrupt. Similarly, the field experiment conducted by Banerjee et al. (2011) is included, but their treatment provided information on politicians' spending discrepancies, which may imply corruption but is not as direct as other types of information provision. The results are not sensitive to the inclusion of these studies (see Figure A.2). The point estimates remain approximately zero percentage points using random effects estimation, and are approximately -1 percentage using fixed effects estimation.

With respect to survey experiments, Chauchard et al. (2017) include two treatments, wealth accumulation and whether the wealth accumulation was is illegal. The effect reported here is the illegal treatment only. This is likely a conservative estimate, as the true effect is a combination of illegality and wealth accumulation. Winters and Weitz-Shapiro (2016) and Weitz-Shapiro and Winters (2017) report results from the same survey experiment. The results are therefore only reported once. The survey experiment in De Figueiredo et al. (2011) is excluded from the analysis as it does not use hypothetical candidates, but instead asks voters if they would have changed their actual voting behavior in response to receiving corruption information. This study has a slightly positive and null finding. The overall results are not sensitive to the inclusion of this estimate (see Figure A.3).

3.4 Results

Survey experiments demonstrate much larger negative treatment effects of providing information about corruption to voters relative to field experiments. In fact, the field-experimental results in Figure 1 reveal a point estimate of approximately zero and suggest that we cannot reject the null hypothesis of no treatment effect. While recognizing the power constraints inherent in such a small sample size, a univariate Shapiro-Wilk test of normality suggests that we cannot reject the null hypothesis that the point estimates are distributed normally around a mean of approximately zero percentage points.

By contrast, corrupt candidates are punished by respondents by approximately 35 percentage points in survey experiments based on fixed effects meta-analysis and 33 percentage points using random effects meta-analysis. Of the 18 survey experiments, only one shows a null effect (Klašnja & Tucker 2013), while all others are negative and significantly different from zero at conventional levels.

Examining all studies together, a test for heterogeneity by type of experiment (field or survey) reveals that up to 66% of the total heterogeneity across studies can be accounted for by a dummy variable for type of experiment (0 = field, 1 = survey). This dummy variable has a significant association with the effectiveness of the information treatment at the 1% significance level. In fact, with this dummy variable included, the overall estimate across studies is approximately 0, while the point estimate of the dummy variable is equal to -0.32.[4] This implies that the predicted treatment effect across experiments is not significantly different from zero when an indicator for type of experiment is included in the model.

[4]Using a mixed effects model with a field experiment moderator.

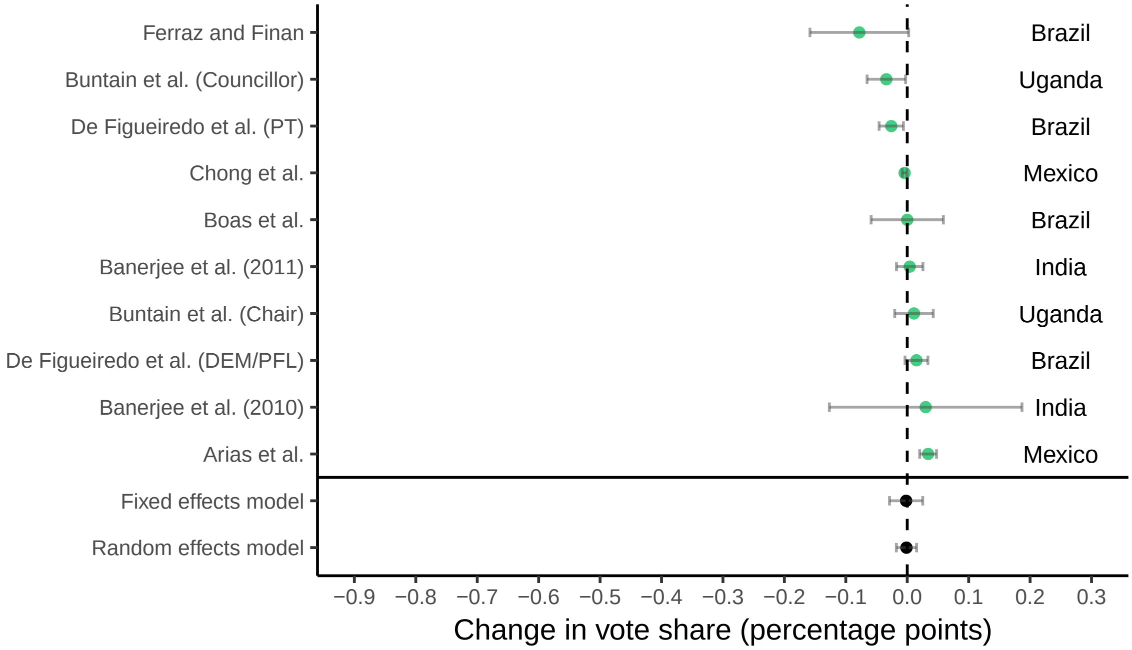

Figure 1: Field experiments: Average treatment effect of corruption information on incumbent vote share and 95% confidence intervals

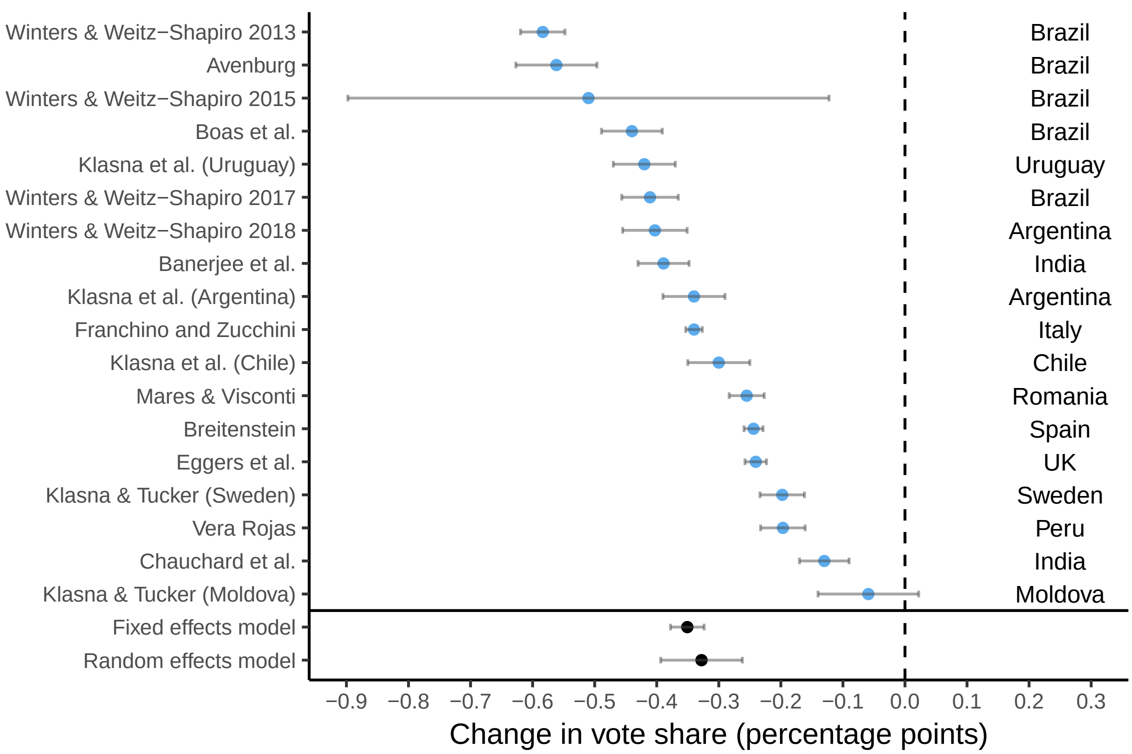

Figure 2: Survey experiments: Average treatment effect of corruption information on incumbent vote share and 95% confidence intervals

4 Exploring the discrepancy

What accounts for the large difference in treatment effects between field and survey experiments? Three possibilities are publication bias, social desirability bias, and the nature of the experimental designs. Null results may be less likely to be published than significant results, particularly in a survey setting. Respondents in survey experiments may also behave in a normatively desirable manner according to the perceived norms of society and/or the researcher. Finally, the experimental design of surveys may not mirror real-world voting decisions. It is possible that more complex factorial designs - such as conjoint experiments - may more successfully approximate real-world settings, and by extension field experiments. However, common methods of analysis of survey experiments may not capture all theoretical quantities of interest.

4.1 Publication bias and p-hacking

Of the ten field experiments I located, only six are published. By contrast, only three of the 18 survey experiment papers remain unpublished, and these are recent drafts. This may reflect that the null results that arise from field experiments are less likely to be published than their survey counterparts with large and highly significant negative treatment effects, even when standard errors are relatively small.

In order to more formally test for publication bias, I first use the p-curve developed in Simonsohn, Nelson, and Simmons (2014a, 2014b) and Simonsohn, Simmons, and Nelson (2015). The p-curve is based on the premise that only "significant" results are typically published, and depicts the distribution of statistically significant p-values for a set of published studies. The shape of the p-curve is indicative of whether or not the results of a set of studies are derived from true effects, or from publication bias. If effect sizes are clustered around 0.05 (i.e. the p-curve is "left skewed"), this may be evidence of p-hacking, indicating that studies with p-values just below 0.05 are "selectively reported." If the p-curve is "right skewed" and there are more low p-values (0.01), this is evidence of true effects.

All significant survey experimental results included in the meta-analysis are significant at the 1% level (making construction of a "curve" with bins of width 0.01 impossible), implying that publication bias likely does not explain the large negative treatment effects in survey experiments.[5] Instead, it appears that the difference in experimental design itself accounts for the difference in the magnitude of treatment effects in field versus survey experiments. For field experiments, there is not a large enough number of published experiments to make the p-curve viable. Only six studies are published, and of these only four are significant at at least the 5% level.

Next, I test for publication bias by examining funnel plot asymmetry. A funnel plot depicts the outcomes from each study on the x-axis and their corresponding standard errors on the y-axis. The chart is overlaid with an inverted triangular confidence interval region (i.e. the "funnel"), which should contain 95% of the studies if there is no bias or between study heterogeneity. If studies with insignificant results remain unpublished the funnel plot may be asymmetric. Both visual inspection and regression tests of funnel plot asymmetry reveal an asymmetric funnel plot when survey and field experiments are grouped together (see Figure A.4 and Table A.3). However, this asymmetry disappears when accounting for heterogeneity by type of experiment, either with the inclusion of a field experiment moderator (dummy) variable or by analyzing field and survey experiments separately (see Figure A.5, Figure A.6, Figure A.7, and Table A.3). Once again, this implies that differences in the experimental design likely account for the difference in the magnitude of treatment effects in field versus survey experiments, rather than publication bias.

4.2 Social desirability bias

A second possible explanation is social desirability bias, in which survey respondents under-report behavior that they believe to be socially undesirable. The respondent may perceive a particular response to be normatively desirable by society as whole, by the researcher(s) conducting the experiment, or both, and respond to the survey in accordance with that norm.

[5] There is also no indication of publication bias at the 1% level.

In the case of corruption, respondents are likely to perceive corruption as both normatively "wrong," as well as harmful to society, the economy, and their own personal well-being. They may therefore be more likely to choose the socially desirable option (no corruption), particularly when the respondent is aware that he or she is being observed by a researcher.

The effect of social desirability bias may be stronger in a survey experiment where there are few downsides to selecting the socially desirable option. A hypothetical vignette has virtually no costs to selecting the socially desirable option, even when moderating variables are included. In a field experiment, however, the cost of changing one's vote may be higher. Voters may have pre-existing opinions of real candidates that make them discount corruption information, or may have strong material and/or ideological incentives to stick with their candidate.

How might we overcome social desirability bias in survey experiments? One option is to eschew hypothetical candidates in favor of real candidates in experiments conducted during the timing of actual elections. Of course, for ethical reasons this likely limits researchers to having actual information regarding the corrupt actions of candidates. A second option is the use of list experiments or experiments which ask about the expected behavior of other individuals in response to new information. List experiments are surprisingly uncommon in corruption experiments (none of the survey experiments included here use this method), but a vote buying experiment in Nicaragua estimated that only 2% of respondents admitted directly to being offered compensation in exchange for their vote, but 24% of respondents admitted to the practice in a list experiment (Gonzalez-Ocantos, De Jonge, Meléndez, Osorio, & Nickerson 2012). A third option, which I turn to next, is the use of more complex factorial designs such as conjoint experiments.

4.3 Do surveys mirror real-world voting decisions?

Previous researchers have noted that even if voters generally find corruption distasteful, the quality of the information provided or positive candidate attributes and policies may outweigh the negative effects of corruption to voters, mitigating the effects of information provision on vote-share.[6] These mitigating factors will naturally arise in a field setting, but may only be salient to respondents if specifically manipulated in a survey setting. A number of survey experiments have therefore added factors other than corruption as mitigating variables.

Corruption accusations can come from a variety of sources, some more credible than others.[7] Multiple studies have therefore attempted to randomize the quality of information provided to voters in order to capture how the electoral penalties vary in response to information quality (Banerjee et al. 2014; Botero et al. 2015; Breitenstein 2019; Mares & Visconti 2019; Weitz-Shapiro & Winters 2017; Winters & Weitz-Shapiro 2018). As expected, higher quality information produces larger negative treatment effects in these experiments (see Figure A.8).

Response to favorable policy stances has been shown to potentially mitigate the impact of corruption to voters. Rundquist et al. (1977) use a survey experiment to show that a candidate's position on the Vietnam War could significantly increase the likelihood of voting for a "corrupt" candidate in the United States. Franchino and Zucchini (2015) similarly show that respondents prefer corrupt but socially and economically progressive candidates to clean but conservative candidates.

Economic benefit has been argued to act as a similar mitigating factor. Klašnja et al. (2017) find evidence that respondents are more forgiving of corruption when it benefits them personally. By contrast, Winters and Weitz-Shapiro (2013) show that respondents punish corrupt politicians, even those with strong records of past performance.

Evidence of co-partisanship as a limiting factor to corruption deterrence is mixed. Anduiza et al. (2013) and Breitenstein (2019) both show that co-partisanship decreases the importance of corruption to Spanish voters in survey experiments. Solaz et al. (2018) induce

[6]See De Vries and Solaz (2017) for a comprehensive overview.

[7]For example, accusations from an independent anti-corruption authority may be deemed more credible than those from an opposition party, and accusations may be deemed less credible than a conviction.

in-group attachment in a lab-experiment of UK subjects, finding that in-group membership reduces sanction of "corrupt" participants. However, Klašnja et al. (2017) find relatively small effects of co-partisanship in Argentina, Chile, and Uruguay, and Rundquist et al. (1977) find null effects in the US in the 1970s. Konstantinidis and Xezonakis (2013) also find that partisanship does not moderate electoral punishment of corruption in a survey experiment in Greece. This evidence suggests that strong partisan effects occur where partisan attachments are strongest.

Yet while the experiments described above capture the tradeoffs voters make between corruption and other electoral factors, they may be measuring a different causal estimand than field experiments. Even if subjects (voters) and treatments (information) are similar, measured outcomes may differ due to the different contexts of the experiments. Consider a voter's choice between two candidates in a field experiment conducted during an election. A candidate is revealed to be corrupt to voters in a treatment group, but not to voters in control. The treated voter can cast a ballot for corrupt candidate A, or candidate B, who is clean. The control voter can cast a ballot for candidate A or candidate B, and has no corruption information. Now consider a survey experiment with a vignette in which the randomized treatment is whether the corrupt actions of a politician are revealed or not. The treated voter can vote for the corrupt candidate A or not, but no challenger exists. Likewise, the control voter can vote for clean candidate A or not, but no challenger exists. Further, the option to abstain from voting does not exist in the survey setting. These differences therefore offer fundamentally different choice sets to voters, altering respondents' potential outcomes and thus capturing different estimands.

4.4 Suggestions for analysis of conjoint experiments

The fact that moderating variables may dampen the salience of corruption to voters has clearly not been lost on previous researchers. However, the meta-analysis indicates that even the inclusion of these moderators does not move point estimates close to the (approximately zero) field setting, in which all of these moderating factors may be salient to the voter. By contrast, conjoint experiments allow researchers to randomize a much larger host of candidate characteristics and may help illuminate the mechanisms that lead to these small effect sizes. In addition, conjoints force respondents to pick between two candidates, better emulating the choice required in an election. Finally, conjoints may minimize social desirability bias as they reduce the probability that the respondent is aware of the researcher's primary experimental manipulation of interest (e.g. corruption).[8]

Researchers have thus far tended to present the results of conjoint experiments as average marginal component effects (AMCEs), then compare the magnitude of these effect sizes. AMCEs represents the unconditional marginal effect of an attribute (e.g. corruption) averaged over all possible values of the other attributes. However, this may or may not be a measure of substantive interest to the researcher. When researchers have strong theories about the conditions that shape voter decision-making, a more appropriate method may be to calculate average marginal effects in order to present predicted probabilities of voting for a candidate under these conditions.[9] For example, in a conjoint experiment including corruption information, this might be interpreted as the probability of voting for a candidate that is both corrupt and possesses other particular feature levels (e.g. party membership or policy positions), marginalizing across all other features in the experiment.

To illustrate this point, I replicate the conjoint experiment conducted in Italy by Franchino and Zucchini (2015) and present both AMCEs and predicted probabilities. Note that I group the "investigated for corruption" and "convicted of corruption" levels into a single "corrupt" attribute in my replications. The predicted probabilities are presented as a function of cor-

[8]This is explicitly mentioned by Hainmueller, Hopkins, and Yamamoto (2014), who argue that conjoint experiments give respondents "various attributes and thus [they] can often find multiple justifications for a given choice." Note, however, that an experiment does not necessarily need to be a conjoint design to have this feature. Conjoint experiments encourage researchers to randomize more attributes and therefore typically contain more complex hypothetical vignettes. However, the same vignette complexity could be achieved without full randomization of these attributes.

[9]This method is utilized by Teele, Kalla, and Rosenbluth (2018) to examine the probability of voting female or male candidates holding other candidate attributes constant, and is discussed in more detail by Leeper, Hobolt, and Tilley (2018).

ruption and two policy positions - tax policy and same sex marriage - for conservative and liberal respondents. The charts therefore show the probability of preferring a profile that has particular levels of tax policy, same sex marriage policy, and corruption, marginalizing across all other features in the experiment.[10]

A casual interpretation of the traditional AMCE plot presented in Figure 3 suggests that it is very unlikely a corrupt candidate would be chosen by a respondent. By contrast, the predicted probabilities plots presented in Figure 4 and Figure 5 show that even for corrupt candidates in the conjoint, the right policy platform can garner over 50% of the predicted hypothetical vote.

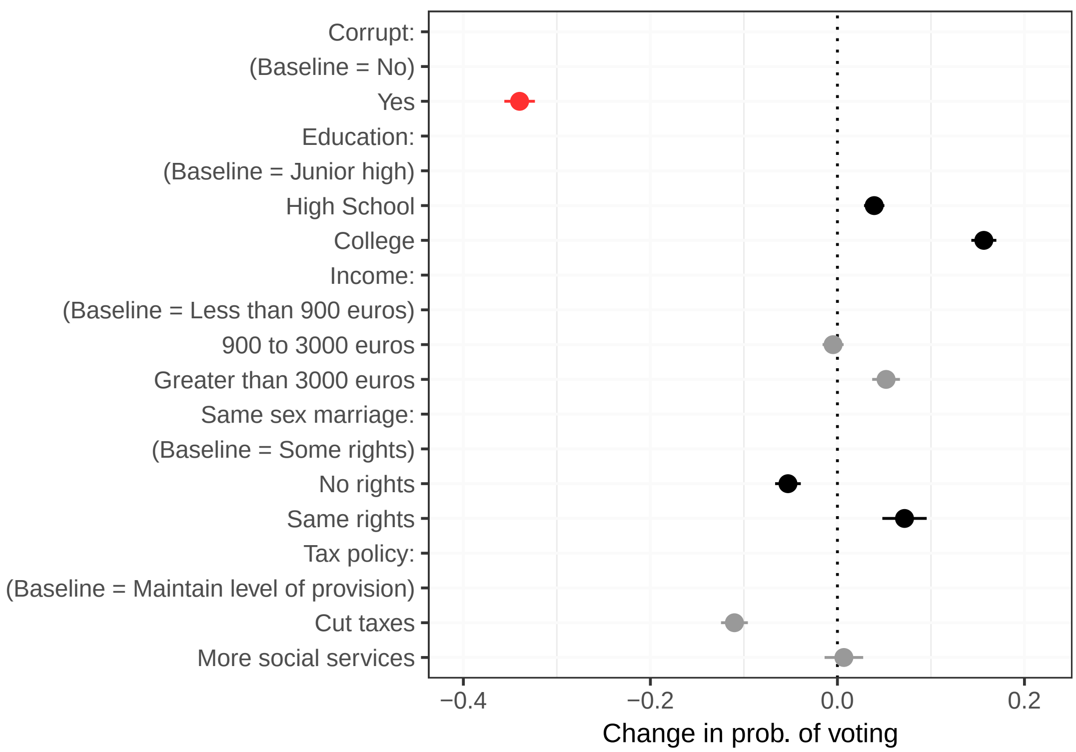

Figure 3: Franchino and Zucchini (2015) conjoint: average marginal component effects

Policy profiles that result in over 50% of voters selecting a "corrupt" candidate may not

[10]Note that the authors correctly conclude that their typical "respondent prefers a corrupt but socially and economically progressive candidate to a clean but conservative one." While I therefore illustrate how predicted probabilities can be used to draw conclusions that may be masked by examination of AMCEs alone, the authors themselves do not make this mistake.

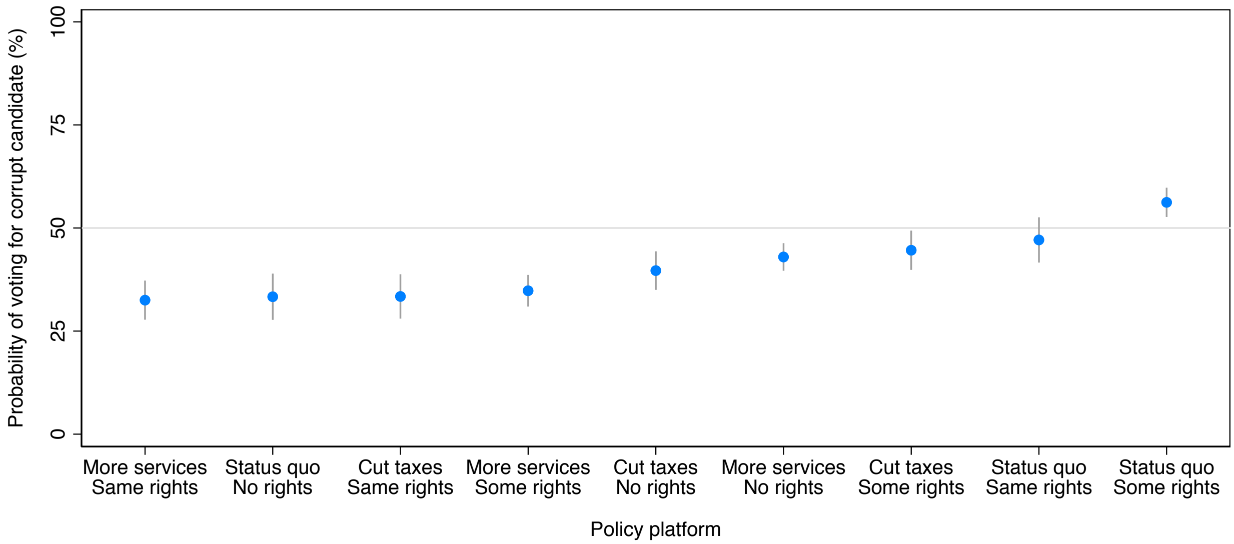

Figure 4: Franchino and Zucchini (2015) conjoint: can policy positions overcome corruption (conservative respondents)?

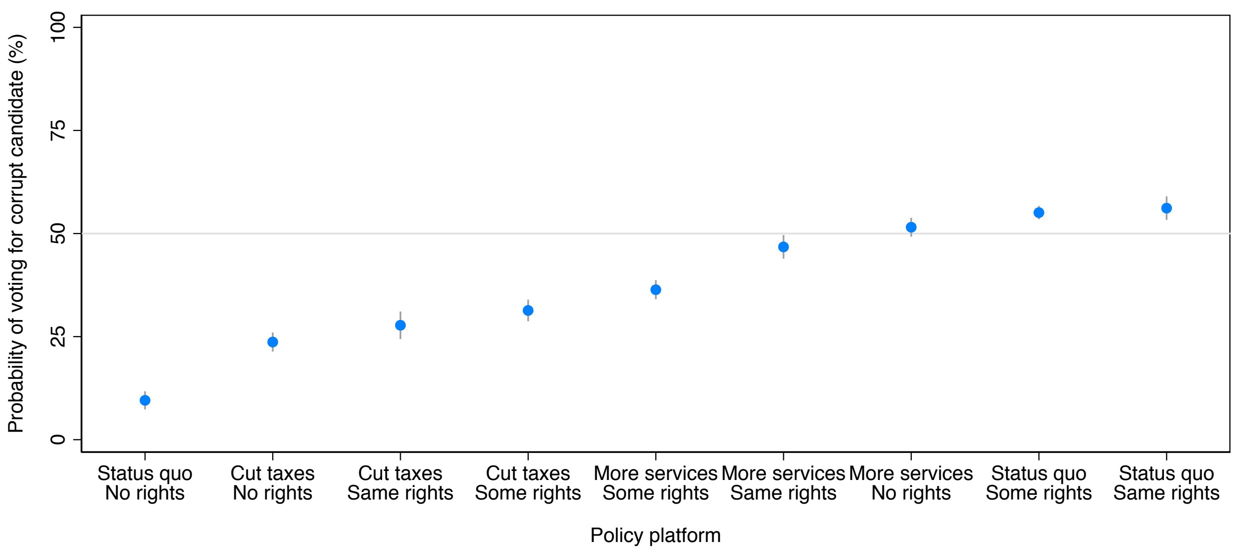

Figure 5: Franchino and Zucchini (2015) conjoint: can policy positions overcome corruption (liberal respondents)?

be outliers in real-world scenarios. Unlike in conjoint experiments, real-world candidates' policy profiles are not selected randomly, but rather represent choices designed to appeal to voters. It may therefore be preferable to analyze conjoint experiments as above, comparing outlier characteristics (e.g. corruption) to realistic candidate profiles rather than fully randomized candidate profiles. For example, in the US context, perhaps the relevant metric of interest would be to look at the impact of corruption on vote choice for a Democratic respondent examining a Democratic candidate who espouses their preferred policy positions and attributes, rather than looking at the magnitude of the corruption AMCE versus each individual policy AMCE.

5 Conclusion

Competitive elections should naturally create a system of accountability, whereby voters expel politicians from office for engaging in malfeasance. However, this mechanism cannot operate without informed voters who are aware of the actions of politicians. In an effort to test whether voters adequately hold politicians accountable for malfeasance, political scientists and economists have turned to experimental methods to test the causal effect of learning about politician corruption on vote choice.

A meta-analytic assessment of these experiments reveals that the conclusions drawn differ drastically depending on whether the experiment was deployed in the field and monitored actual vote choice, versus hypothetical vote choice in an online setting. Across field experiments, the aggregate treatment effect of providing information about corruption on vote share is approximately zero. By contrast, in survey experiments corrupt candidates are punished by respondents by approximately 33-35 percentage points.

I explore three possible explanations that may explain this discrepancy: publication bias, social desirability bias, and the nature of the experimental designs. I do not find systematic evidence of publication bias. Survey experiments may be capturing strong anti-corruption norms rather than realistic voter behavior, pushing voters to select the clean candidate in hypothetical vignettes where few candidate traits are known. Field and survey experiments are likely also measuring different causal estimands, as field experiments possess the option of voting for alternative candidates or abstaining.

High-dimension factorial designs such as conjoint experiments may alleviate some of these issues. However, it may be preferable to analyze vote-choice conjoint experiments by comparing the probability of voting for a candidate with outlier characteristics such as corruption to the probability of voting for a realistic candidate without this characteristic, rather than examining differences in AMCEs across fully randomized candidate profiles, since these are untenable and so tell us little about real-world electoral choices.

These findings suggest that while vote-choice survey experiments may provide information on the directionality of informational treatments in idealized hypothetical scenarios, the point estimates they provide may not be representative of real-world voting behavior. More generally, researchers should exercise caution when interpreting actions taken in hypothetical vignettes as indicative of real world behavior such as voting.

References

Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comparative Political Studies*, *46*(12), 1664–1692.

Arias, E., Larreguy, H., Marshall, J., & Querubin, P. (2018). *Priors rule: When do malfeasance revelations help or hurt incumbent parties?* (Tech. Rep.). National Bureau of Economic Research.

Arias, E., Larreguy, H., Marshall, J., & Querubin, P. (2019). Information provision, voter coordination, and electoral accountability: Evidence from mexican social networks. *American Political Science Review*.

Arvate, P., & Mittlaender, S. (2017). Condemning corruption while condoning inefficiency: an experimental investigation into voting behavior. *Public Choice*, *172*(3-4), 399–419.

Avenburg, A. (2016). *Corruption and electoral accountability in brazil* (Unpublished doctoral dissertation).

Azfar, O., & Nelson, W. R. (2007). Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice*, *130*(3-4), 471–493.

Banerjee, A., Green, D., Green, J., & Pande, R. (2010). Can voters be primed to choose better legislators? experimental evidence from rural india. In *Presented and the political economics seminar, stanford university.*

Banerjee, A., Green, D. P., McManus, J., & Pande, R. (2014). Are poor voters indifferent to whether elected leaders are criminal or corrupt? a vignette experiment in rural india. *Political Communication*, *31*(3), 391–407.

Banerjee, A., Kumar, S., Pande, R., & Su, F. (2011). Do informed voters make better choices? experimental evidence from urban india. *Unpublished manuscript*.

Boas, T. C., Hidalgo, F. D., & Melo, M. A. (2018). Norms versus action: Why voters fail to sanction malfeasance in brazil. *American Journal of Political Science*.

Botero, S., Cornejo, R. C., Gamboa, L., Pavao, N., & Nickerson, D. W. (2015). Says who? an experiment on allegations of corruption and credibility of sources. *Political Research Quarterly*, *68*(3), 493–504.

Breitenstein, S. (2019). Choosing the crook: A conjoint experiment on voting for corrupt politicians. *Research & Politics*, *6*(1), 2053168019832230.

Buntaine, M. T., Jablonski, R., Nielson, D. L., & Pickering, P. M. (2018). Sms texts on corruption help ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences*, *115*(26), 6668–6673.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list. *Available at SSRN 1977749*.

Chauchard, S., Klasnja, M., & Harish, S. (2017). The limited impact of information on political accountability: An experiment on financial disclosures in india. *Unpublished manuscript). Dartmouth College, Hanover, NH*.

Chong, A., De La O, A. L., Karlan, D., & Wantchekon, L. (2014). Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics*, *77*(1), 55–71.

Coppock, A., & Green, D. P. (2015). Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, *3*(1), 113–131.

De Figueiredo, M. F., Hidalgo, F. D., & Kasahara, Y. (2011). When do voters punish corrupt politicians? experimental evidence from brazil. *Unpublished manuscript, UC Berkeley*.

De Vries, C. E., & Solaz, H. (2017). The electoral consequences of corruption. *Annual Review of Political Science*, *20*, 391–408.

Eggers, A. C., Vivyan, N., & Wagner, M. (2018). Corruption, accountability, and gender: do female politicians face higher standards in public life? *The Journal of Politics*, *80*(1), 321–326.

Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics*, *123*(2), 703–745.

Franchino, F., & Zucchini, F. (2015). Voting in a multi-dimensional space: a conjoint analysis employing valence and ideology attributes of candidates. *Political Science Research and Methods*, *3*(2), 221–241.

Gonzalez-Ocantos, E., De Jonge, C. K., Meléndez, C., Osorio, J., & Nickerson, D. W. (2012). Vote buying and social desirability bias: Experimental evidence from nicaragua. *American Journal of Political Science*, *56*(1), 202–217.

Gray, C. W., & Kaufman, D. (1998). Corruption and development.

Green, D. P., Zelizer, A., Kirby, D., et al. (2018). Publicizing scandal: Results from five field experiments. *Quarterly Journal of Political Science*, *13*(3), 237–261.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, *22*(1), 1–30.

Humphreys, M., & Weinstein, J. (2012). Policing politicians: citizen empowerment and political accountability in uganda preliminary analysis. *Columbia Universities. Unpublished manuscript*.

Klašnja, M., Lupu, N., & Tucker, J. A. (2017). When do voters sanction corrupt politicians?

Klašnja, M., & Tucker, J. A. (2013). The economy, corruption, and the vote: Evidence from experiments in sweden and moldova. *Electoral Studies*, *32*(3), 536–543.

Kolstad, I., & Wiig, A. (2009). Is transparency the key to reducing corruption in resource-rich countries? *World development*, *37*(3), 521–532.

Konstantinidis, I., & Xezonakis, G. (2013). Sources of tolerance towards corrupted politicians in greece: The role of trade offs and individual benefits. *Crime, Law and Social Change*, *60*(5), 549–563.

Leeper, T. J., Hobolt, S. B., & Tilley, J. (2018). Measuring subgroup preferences in conjoint experiments.

Mares, I., & Visconti, G. (2019). Voting for the lesser evil: Evidence from a conjoint experiment in romania. *Political Science Research and Methods*.

Muñoz, J., Anduiza, E., & Gallego, A. (2012). Why do voters forgive corrupt politicians? cynicism, noise and implicit exchange. In *International political science association conference. madrid, spain.*

Rose-Ackerman, S., & Palifka, B. J. (2016). *Corruption and government: Causes, consequences, and reform*. Cambridge university press.

Rundquist, B. S., Strom, G. S., & Peters, J. G. (1977). Corrupt politicians and their electoral support: some experimental observations. *American Political Science Review*, *71*(3), 954–963.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666–681.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to ulrich and miller (2015).

Solaz, H., De Vries, C. E., & de Geus, R. A. (2018). In-group loyalty and the punishment of corruption. *Comparative Political Studies*, 0010414018797951.

Teele, D. L., Kalla, J., & Rosenbluth, F. (2018). The ties that double bind: Social roles and women's underrepresentation in politics. *American Political Science Review*, *112*(3), 525–541.

Vera Rojas, S. B. (2017). The heterogeneous effects of corruption: Experimental evidence from peru. *Manuscript, University of Pittsburgh*.

Weitz-Shapiro, R., & Winters, M. S. (2017). Can citizens discern? information credibility, political sophistication, and the punishment of corruption in brazil. *The Journal of Politics*, *79*(1), 60–74.

Weschle, S. (2016). Punishing personal and electoral corruption: Experimental evidence from india. *Research & Politics*, *3*(2), 2053168016645136.

Winters, M. S., & Weitz-Shapiro, R. (2013). Lacking information or condoning corruption: When do voters support corrupt politicians? *Comparative Politics*, *45*(4), 418–436.

Winters, M. S., & Weitz-Shapiro, R. (2015). Political corruption and partisan engagement: evidence from brazil. *Journal of Politics in Latin America*, *7*(1), 45–81.

Winters, M. S., & Weitz-Shapiro, R. (2016). Who?s in charge here? direct and indirect accusations and voter punishment of corruption. *Political Research Quarterly*, *69*(2), 207–219.

Winters, M. S., & Weitz-Shapiro, R. (2018). Information credibility and responses to corruption: a replication and extension in argentina. *Political Science Research and Methods*, 1–9.

A Appendix

A.1 *Lab experiments*

Table A.1: Lab experiments

Study	Country	ATE
Arvate and Mittlaender (2017)	Brazil	Negative
Azfar and Nelson (2007)	USA	Negative
Rundquist et al. (1977)[1]	USA	Negative
Solaz et al. (2018)	UK	Negative

[1] The candidate is always corrupt in the Rundquist et al. (1977) experiment. A "corruption" point estimate is therefore not provided in the coefficient plot below.

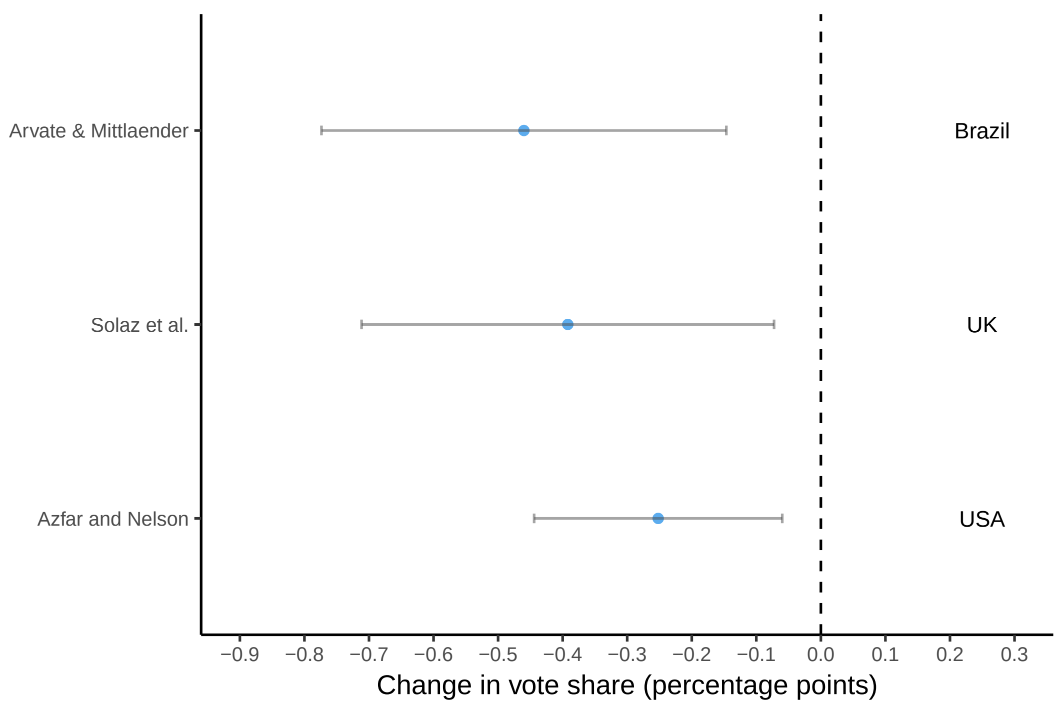

Figure A.1: Lab experiments: Average treatment effect of corruption information on vote share

.

A.2 Excluded studies

Table A.2: Excluded experiments

Study	Type	Reason for exclusion
Anduiza et al. (2013)	Survey	Lack of no-corruption control group
Botero et al. (2015)	Survey	Lack of no-corruption control group
De Figueiredo et al. (2011)	Survey	Outcome is hypothetically changing actual vote
Konstantinidis and Xezonakis (2013)	Survey	Lack of no-corruption control group
Muñoz et al. (2012)	Survey	Lack of no-corruption control group
Rundquist et al. (1977)	Lab	Lack of no-corruption control group
Weitz-Shapiro and Winters (2017)	Survey	Data identical to Winters and Weitz-Shapiro (2016)
Weschle (2016)	Survey	Lack of no-corruption control group

A.3 Robustness checks

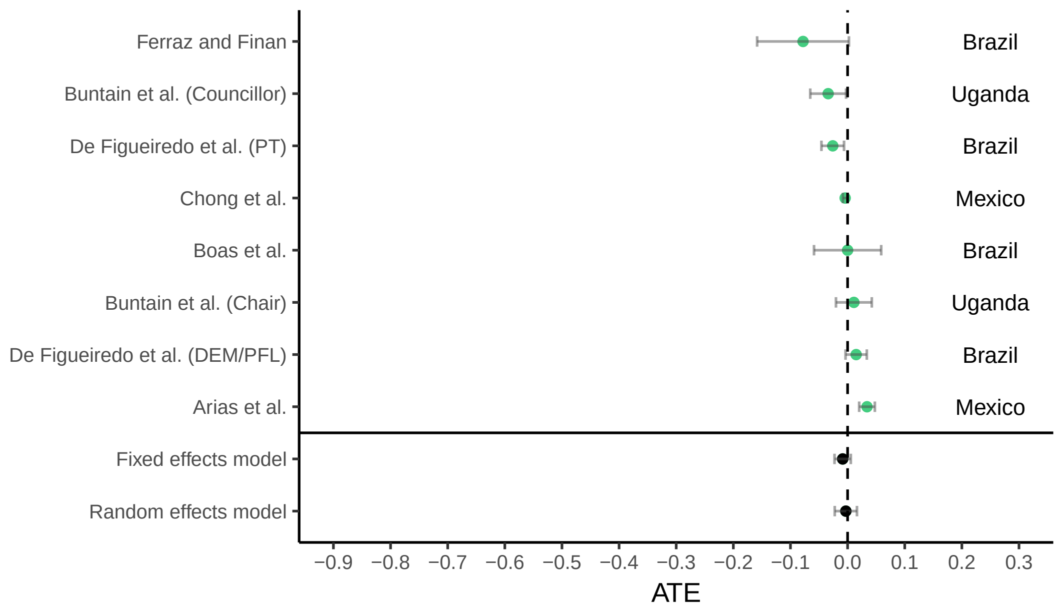

Figure A.2: Field experiments: Average treatment effect of corruption information on incumbent vote share (excluding Banerjee et al. (2010) and Banerjee et al. (2011))

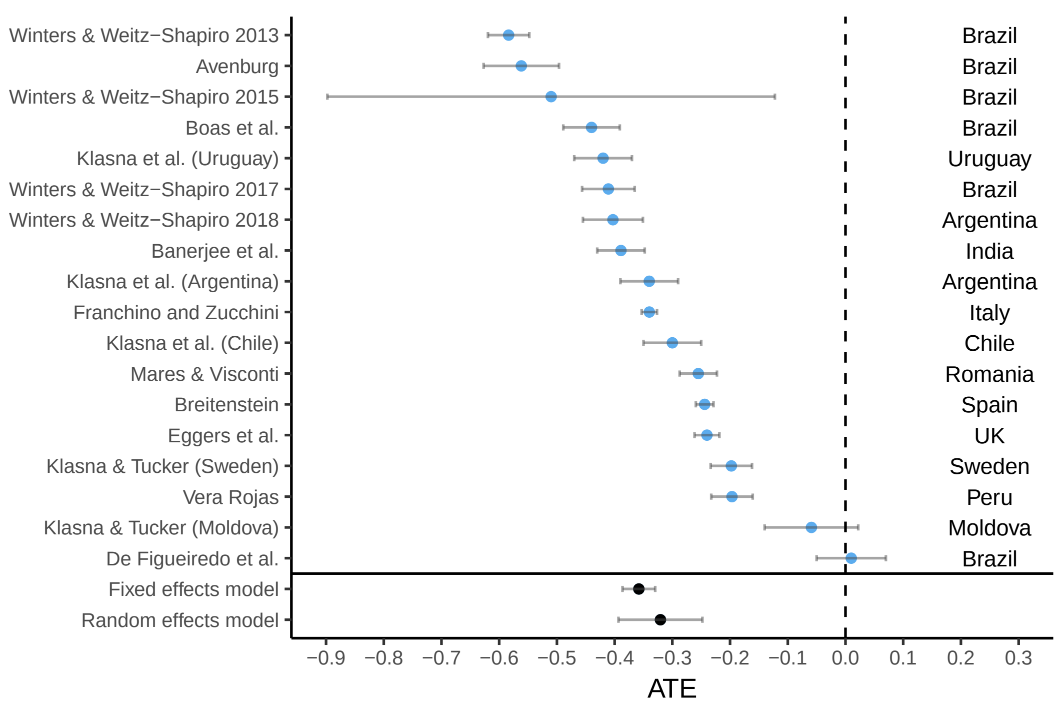

Figure A.3: Survey experiments: Average treatment effect of corruption information on incumbent vote share (including De Figueiredo et al. (2011))

.

A.4 Publication bias

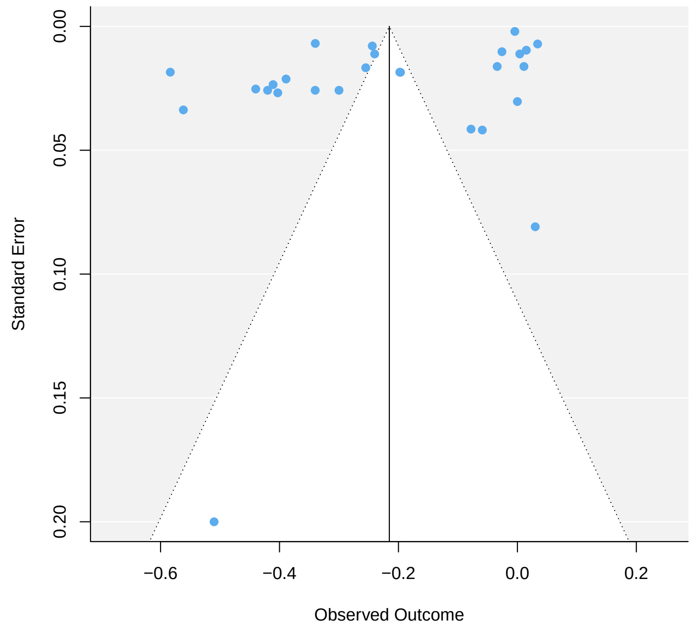

Figure A.4: Funnel plot: all experiments

.

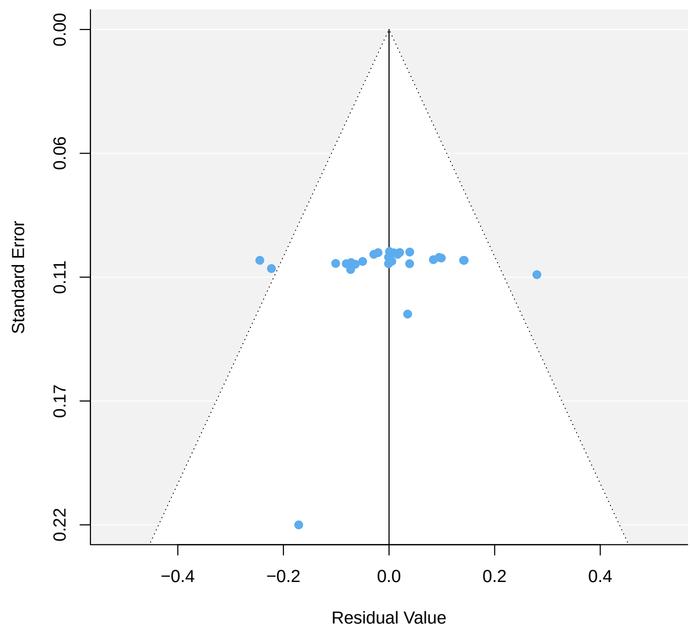

Figure A.5: Funnel plot: all experiments with field experiment moderator

.

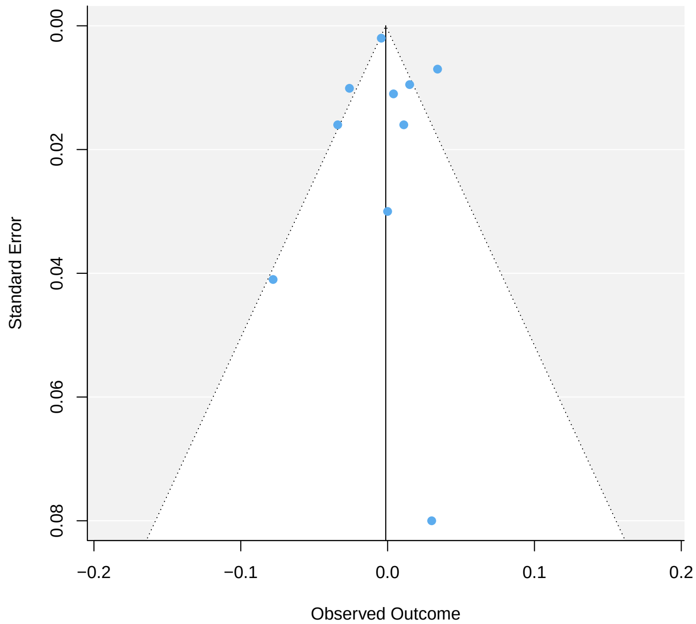

Figure A.6: Funnel plot: field experiments

.

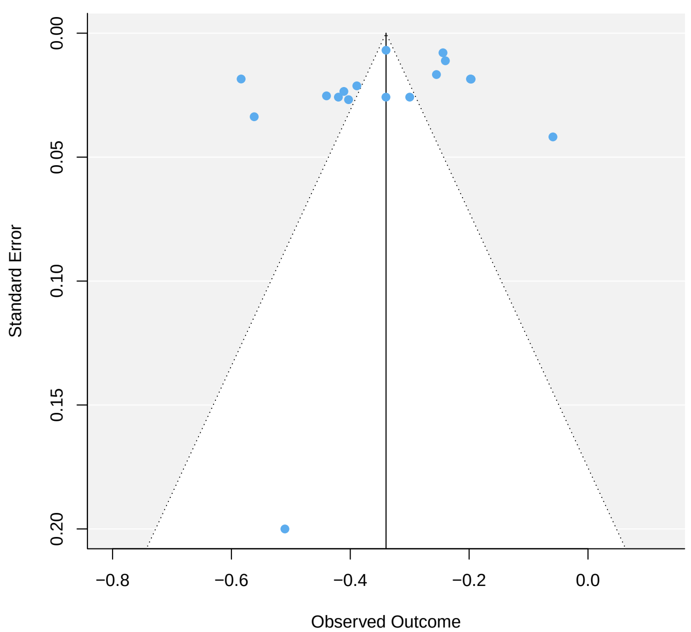

Figure A.7: Funnel plot: survey experiments

.

Table A.3: Regression tests for funnel plot asymmetry

Studies included	p-value
All	0.0016
All with moderator	0.4512
Field	0.8403
Survey	0.3159

A.5 Information quality

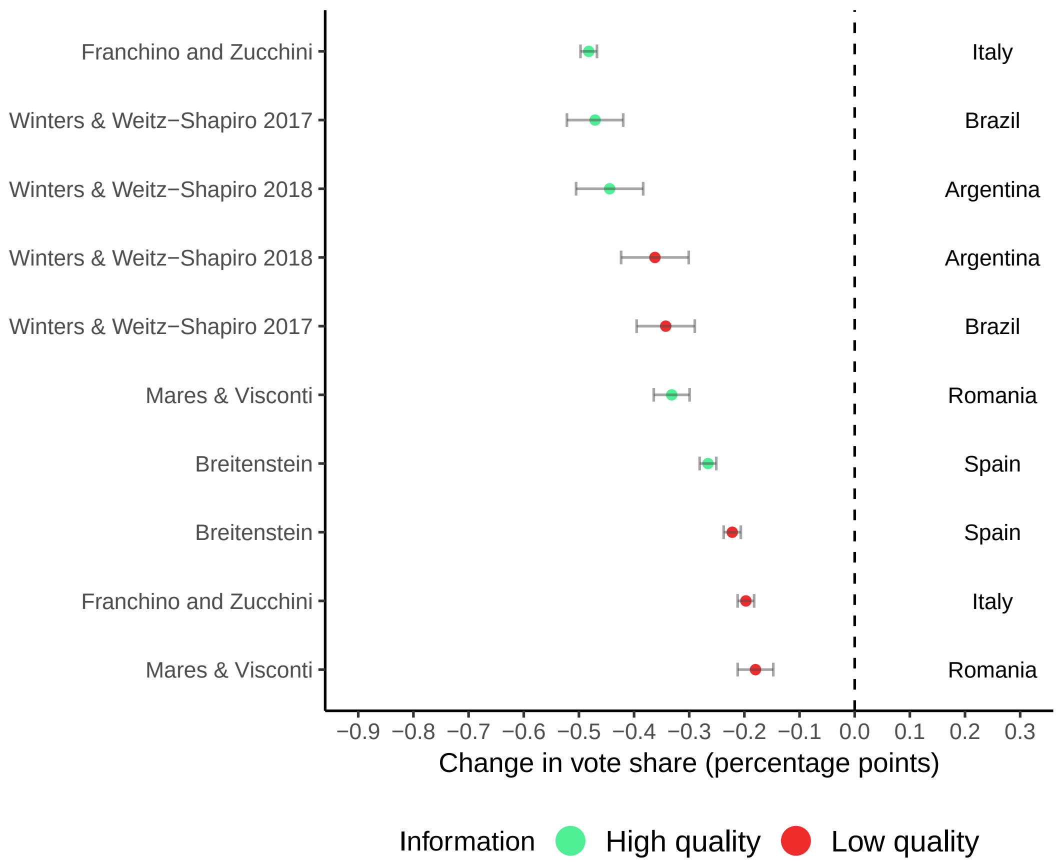

Figure A.8: Survey experiments by information quality: Average treatment effect of corruption information on vote share

.