Journal of Development Economics 131 (2017) 1–14

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

# Temptation in vote-selling: Evidence from a field experiment in the Philippines

Allen Hicken [a], Stephen Leider [b], Nico Ravanilla [c], Dean Yang [d,*]

[a] Department of Political Science, University of Michigan, USA
[b] Stephen M. Ross School of Business, University of Michigan, USA
[c] School of Global Policy & Strategy, University of California San Diego, USA
[d] Department of Economics and Gerald R. Ford School of Public Policy, University of Michigan; BREAD; and NBER, USA

ARTICLE INFO

*Keywords:*
Vote-selling
Vote-buying
Temptation
Self-control
Commitment
Elections
Political economy
Philippines

ABSTRACT

We report the results of a randomized field experiment in the Philippines on the effects of two common anti-vote-selling strategies involving eliciting promises from voters. An invitation to promise not to vote-sell is taken up by most respondents, reduces vote-selling, and has a larger effect in races with smaller vote-buying payments. The treatment reduces vote-selling in the smallest-stakes election by 10.9 percentage points. Inviting voters to promise to “vote your conscience” despite accepting money is significantly less effective. The results are consistent with a behavioral model in which voters are only partially sophisticated about their vote-selling temptation.

## 1. Introduction

Vote-buying and vote-selling are pervasive phenomena in many developing democracies. While there is some debate about the consequences of the buying and selling of votes, there is a consensus that transactional electoral politics brings with it a host of costs. For example, vote-buying and other forms of clientelism can undermine or even reverse the standard accountability relationship that is central to democracy (Hicken, 2011; Kitschelt et al., 2010; Lyne, 2007; Stokes, 2005; Stokes et al., 2013). Vote-buying also hampers the development of and trust in the political institutions necessary for democratic development and consolidation (Desposato, 2007; Graziano, 1973; Kitschelt et al., 2010; Lyne, 2007; Stokes, 2005). Finally, vote-buying and other forms of clientelism are associated with larger public deficits and public sector inefficiencies (Hicken and Simmons, 2008; Keefer, 2006, 2007), and higher levels of corruption (Kitschelt and Wilkinson, 2007; Kitschelt et al., 2010; Keefer, 2007).

Because of these potential inimical effects, governments, NGOs, and international donors have directed significant attention and resources towards combating vote-buying and vote-selling. Some strategies focus on the demand side of the equation—making it more difficult for politicians (or vote-*buyers*) to offer money in exchange for a vote. However, such strategies often fall victim to poor implementation and enforcement. As a result, a major focus of anti-vote-buying efforts has been on vote-*sellers*. Whether organized by governmental election commissions, or by concerned NGOs, campaigns to reduce the supply of votes available for purchase are common worldwide. Voter-focused campaigns against vote-selling tend to fall into two categories. The first type of campaign urges voters to avoid taking vote-buying payments at all. Voters may be asked to make promises or sign pledges to simply eschew taking money from politicians or their agents prior to elections. A second common approach seeks to subvert vote-buying by encouraging voters to take the money being offered, but nonetheless “vote their conscience.” For example, Cardinal Sin, Archbishop of Manila, famously advised voters to “take the bait, not the hook” (Schaffer, 2005).[1]

Motivated by both the negative consequences of transactional electoral politics, and by the prevalence of anti-vote-selling efforts, in this paper, we seek to deepen our understanding of the political economy and psychology of individual vote-selling decisions. A number of questions are of general interest. What is the efficacy of anti-vote-selling campaigns? Can simple promises—such as the ones elicited from voters in anti-vote-selling campaigns—affect vote-selling behavior? If so, why might voters make such promises? Does the impact of promises differ by type of promise (e.g. “I won’t take money” vs. “I’ll take money, but

* Corresponding author.
*E-mail addresses:* ahicken@umich.edu (A. Hicken), leider@umich.edu (S. Leider), nravanilla@ucsd.edu (N. Ravanilla), deanyang@umich.edu (D. Yang).

[1] For examples of both types of campaigns, see Callahan (2000); Guiang (2013); Geronimo (2013); Schaffer (2005).

https://doi.org/10.1016/j.jdeveco.2017.10.012
Received 22 August 2016; Received in revised form 21 September 2017; Accepted 23 October 2017
Available online 28 October 2017
0304-3878/Published by Elsevier B.V.

vote my conscience")? Might some types of promises actually *increase* the incidence of vote-selling?

We ask these questions in the context of a randomized controlled trial of an anti-vote-selling intervention in Sorsogon City, Philippines. We randomly assigned voters to a control group or to one of two treatment groups, and examine impacts on a proxy for vote-selling based on self-reports of respondents' voting behavior (from a post-election survey).

The treatments we examine were designed to mirror the types of promises elicited in anti-vote-selling campaigns. In the Promise 1 treatment, we invited voters to promise not to take vote-buying payments at all. In the Promise 2 treatment, we invited voters to promise that if they did take vote-buying payments, they would nevertheless vote their conscience.

We estimate the impacts of the promise treatments on a proxy for vote-selling: "vote-switching". This outcome variable is based on comparison of respondent self-reports in a pre-election survey and a post-election survey. In the pre-election survey, fielded some weeks before the election, we ask respondents for favorability ratings of candidates. Then, in the post-election survey, we ask respondents who they actually voted for. We define vote-switching as reporting that one voted for a candidate who was not rated as one's favorite in the pre-election survey.[2] We look at vote-switching in three local races: the elections for mayor, vice-mayor and city council. While examining vote-switching is an indirect way of getting at vote-selling, vote-switching is self-reported, which raises concerns about social desirability bias: respondents could respond to the promise treatments by falsely maintaining consistency between their pre-election ratings and their post-election voting reports. Such biased reporting could lead us to spuriously find that the promise treatments reduce vote-switching.[3]

Our results provide a reasonably strong indication that social desirability bias is not a significant concern in our setting. Support for this claim comes from comparisons of the treatment effects of Promise 1 ("Don't take the money") on vote-switching across electoral races.[4] One would expect social desirability bias to be constant across electoral races, or increasing in the importance of the race. In our setting, if there were only social desirability bias and no "true" treatment effects, this would mean that we should find larger (negative) treatment effects for the mayor and vice-mayor races, compared to the city council race. As it turns out, we find the opposite to be true: the Promise 1 treatment effects on vote-switching, while negative, are very close to zero in the two most important electoral races that we examine (the elections for mayor and vice-mayor.) By contrast, we find much larger negative effects on vote-switching in the city council election, the least important of the races. We conclude from this comparison that our treatment effect estimates are minimally biased (if at all) by intentional misreporting.

We estimate that the Promise 1 treatment reduced vote-switching (and therefore vote-selling) in the race involving smaller vote buying payments (the city council race) by 10.9 percentage points. Compared to the city-council vote-switching rate of 47.1 percent in the control group, this is a large effect, given that vote-switching can occur for reasons other than vote-selling. As mentioned previously, the impacts of the Promise 1 treatment on vote-switching in the more important races (mayor and vice-mayor) are close to zero and are not statistically significant.

We also conduct statistical tests of pairwise differences in treatment effects across promise types (within electoral races), and across races (within promise types). We find that the Promise 1 treatment has a more negative effect on vote-switching than does the Promise 2 treatment. We also find that the promise treatments reduce vote-switching more for races with lower vote-buying payments (the city council race) than in the higher-money races (the mayor and vice-mayor races).

To help explain this pattern of heterogeneity in impacts across promises and electoral races, we developed, *ex-post*, a behavioral model of transactional electoral politics. We model selling one's vote as a temptation good: it creates positive utility for the future self at the moment of voting, but not for past selves who anticipate the sale of the vote. In addition, voters can make promises in advance of elections regarding whether or not they will sell their votes, and gain (lose) utility when they keep (break) such promises. We also allow for the possibility that voters may not be fully sophisticated about their vote-selling temptation. Specifically, when deciding whether to accept a gift from a candidate, they may underestimate how much utility the future self will gain from voting for the candidate who provided the gift (said another way, they underestimate the impact of accepting vote-buying payments today on their propensity to vote for the vote-buying candidate in the future.) The model also implies that voters who are at least partially sophisticated about their vote-selling temptation can use promises not to take money from candidates at all as a commitment device.

The pattern of our empirical results is consistent with the case of the model in which voters are partially aware of their vote-selling temptation (neither fully aware nor fully naïve of it). In the model, the worse performance of Promise 2 comes from respondents who would not have accepted money if they had been in the control group, but who (incorrectly) believe they can accept money without changing their vote due to making the promise. By contrast, a fully sophisticated voter correctly anticipates his temptation, so would not make this mistake. Fully naïve voters would not increase their uptake of money offers due to the promise treatments, since they would accept money in the control treatment as well.

Our research is related to work on electoral malpractices more generally. Existing research has established, via natural experiments in a variety of contexts, that electoral malpractices have material influence on election outcomes (Golden and Tiwari, 2009; Acemoglu et al., 2009; Baland and Robinson, 2008; Golden et al., 2014). On the specific topic of vote-selling, research has shown it to be more prevalent among poor voters (Scott, 1969; Stokes, 2005; Blaydes, 2006; Bratton, 2008), and that parties, candidates and brokers are often strategic regarding which populations they target for vote-buying (Stokes et al., 2013). Khemani (2013) finds that the extent of vote-buying is negatively correlated with public health service delivery across municipalities in one Philippine province. Banerjee et al. (2011) find, in the context of a randomized controlled trial in urban India, that provision of "report cards" comparing electoral candidates reduces vote buying and leads to higher vote shares for higher-quality candidates. Finan and Schechter (2012) find that vote-buying payments in rural Paraguay are targeted to "reciprocal" individuals (as measured in an artefactual field experiment), suggesting that vote-buying exploits informal norms of reciprocity. Vicente (2014) conducted a randomized controlled trial of an anti-vote-selling intervention, finding that it raised the vote share of incumbents, consistent with challengers' use of vote-buying to overcome incumbency advantages. Cruz et al. (2015) find in the Philippines that provision of information to voters on candidates' spending priori-

[2] Individuals can be "vote-switchers" for many reasons aside from vote-selling (such as learning new information about candidates), but, given random assignment, the promise treatments should only affect vote-switching via changes in vote-selling.

[3] Our study faces challenges similar to those faced by survey experiments in political science (Gaines et al., 2007). In addition to social desirability bias, concerns related to external validity (the extent to which survey self-reports actually reflect real-world behavior) are central (Barabas and Jerit, 2010).

[4] As we discuss further below, analogous comparisons across races for Promise 2 treatment effects are not as revealing of the extent of social desirability bias because Promise 2, in principle, can actually raise vote-switching.

ties led those voters to be targeted for vote-buying.[5]

In its focus on the real-world impact of promises, this paper is also related to recent work from behavioral psychology and economics that shows that promises and other informal agreements can substantially change behavior and lead to more socially efficient outcomes by changing social norms (Charness and Dufwenberg, 2006; Vanberg, 2008; Kessler and Leider, 2012; Krupka et al., 2013). Shu et al. (2012) show that the form of promise elicitation affects honesty in reporting of information in auto insurance applications. We also have a clear connection to research on temptation goods (Banerjee and Mullainathan, 2010; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2001) and on self-control problems (Giné et al., Forthcoming; Laibson, 1997; Ashraf et al., 2006; Duflo et al., 2011; Kaur et al., 2017).

## 2. Context and overview of vote-buying

The experiment was conducted in Sorsogon City, Sorsogon Province, Philippines. Sorsogon Province is located at the southern tip of Luzon island, roughly 12 h by road from the national capital, Manila. Sorsogon City, with a population of roughly 150,000, is the provincial capital, and is slightly below the median across Philippine municipalities in terms of economic development. With a municipal poverty rate of 35%, it is slightly worse than the median (the 45th percentile, to be exact) poverty rate among Philippine municipalities.[6]

We study voting behaviors in the 2013 elections for Sorsogon City municipal positions (mayor, vice-mayor, and city council). The mayoral and vice-mayoral elections are the more important races at the local level. The mayor is the chief executive of the city government, and among its many powers (see Local Government Code of the Philippines 1991) is to direct the formulation of the city government plan, issue executive orders, and represent the city in all its business transactions and sign on its behalf all bonds, contracts, and obligations. The vice-mayor is the presiding officer of the city council and signs all warrants drawn on the municipal treasury for all expenditures appropriated for the operation of the council. The vice-mayor also appoints all officers and employees of the council. The city council has the legislative power, including the power to approve ordinances and pass resolutions necessary for an efficient and effective city government, as well as the power to approve or veto the annual and supplemental budgets of the city government.

Mayors and vice-mayors do not run in pairs, and winners sometimes come from different parties (often yielding a divided executive). City council members are elected from a single (district) constituency, using block voting: voters may vote for up to four councilors, with the top four vote-getters in a district being awarded council seats. Both the split-ticket mayoral and vice-mayoral race and the block vote system for city council seats tend to undermine the value of party affiliation (or running in a single ticket) and encourage individual candidates to develop personalized networks of support (Hicken, 2009).

As in many other parts of the Philippines, vote-buying is widespread in our study location, with nearly every candidate participating, from Congress down to council candidates. We define vote-buying as the offer of resources by political campaigns to individuals or households in order to persuade them to vote for a particular candidate. This definition is consistent with the definitions elsewhere in the literature (e.g. Stokes et al., 2013; Vicente, 2014).

We provide details about the logistics of vote buying in the Online Appendix, but here we present a brief summary. Most vote-buying in Sorsogon City occurs in the week leading up to election day.[7] Using voter lists each campaign has developed, candidate representatives approach households directly, offering money or goods in exchange for their vote. Based on observations of our project field staff, vote-buying payments differed substantially across races. In the mayor and vice-mayor races, payments typically amounted to 250 to 500 Philippine pesos, while those for city council were in the range of 20–100 pesos.[8]

Vote buying is done systematically and strategically. Typically, each voter in a household will be offered a packet with their name on it, and campaigns track who accepted and who did not. Candidates may also engage in a second round of vote buying if they learn that a challenger is offering more money than they are. Campaigns seek to ensure that voters clearly associate the gift with their candidate. For example, the candidate's flyer may be stapled to packages of food handed out to voters or cash may be attached to flyer or letter from the candidate. Most commonly, candidates distribute money attached to a sample ballot, and encourage voters to take the ballots with them to the polls as a guide. The sample ballot includes not just the candidate's name, but also allied candidates from other races up and down the ticket.[9] For further background, including images of sample ballots, please see the Online Appendix.

## 3. Experimental design and data collection

We implemented a randomized controlled trial of treatments encouraging individual voters not to sell their votes. Study participants were registered voters in Sorsogon City. Participants were selected from the Certified List of Voters that we obtained from the Commission on Elections (COMELEC). The list included the name, address, date of birth, gender, and the assigned polling precinct of each of Sorsogon City's 84,284 registered voters.[10] From this list, we randomly selected 900 primary targeted respondents and 900 alternates.

Prior to fielding the baseline survey and intervention, primary respondents and alternates were randomly assigned to the control or treatment groups. One-third of individuals were randomly assigned to the control group, one-third to the Promise 1 treatment, and one-third to the Promise 2 treatment.

### *3.1. Baseline survey and voter educational video*

The baseline survey and treatments were administered prior to the May 13, 2013 elections for Sorsogon City mayor, vice-mayor, and city council. A local team of enumerators administered the baseline survey, treatment interventions, and the endline survey. Surveys were administered on a hand-held device (an iPad) using an offline survey app (iSurvey). The baseline survey was fielded from April 17 to May 8, 2013 (5–26 days prior to the election).

Enumerators located primary respondents at their residential addresses, invited them to participate in the research study using a

[5] There is of course a larger related literature on voter decision-making, separately from vote-selling or -buying. Olken and Pande (2011) survey recent research (using experimental and observational methods) demonstrating that voter behavior is highly malleable, and information provision in the context of elections can improve electoral accountability in developing country democracies. Recent studies of note include Wantchekon (2003); Ferraz and Finan (2008); Banerjee et al. (2012); Chong et al. (2011); Gine and Mansuri (2012); Beaman et al. (2009).

[6] Poverty rates are from 2003. The Philippines' overall poverty incidence is 29% (National Statistical Coordination Board 2009).

[7] For some candidates vote buying may be the culmination of long-term efforts at cultivating voter loyalty via constituency service or other strategies.

[8] According to the Commission on Elections (COMELEC), they received reports of vote buying from all over the country during the May 2013 elections, with reported amounts ranging from P200 to P5000 (Flores et al., 2013). See also (Quijano, 2013).

[9] While candidates may encourage voters to vote up and down the same ticket, typically only money from one candidate at a time is offered to voters. This is done explicitly to avoid free-riding and maximize the chances that voters will assign credit for the money to the correct candidate. In the comparatively rare case where candidates from different offices coordinate to distribute money together voters will typically receive two separate envelopes/sample ballots, with money attached—one from each candidate.

[10] The registration deadline for the May 2013 elections was October 2012, so this list was the complete list of registered voters for our election of interest.

recruitment script (see Online Appendix B), and obtained consent to participate in the study. When a primary respondent could not be interviewed due to out-migration, refusal, or being deceased, the enumerator sought to interview an alternate respondent with the same treatment assignment.[11] Following this procedure, we generated a sample of 883 respondents, just slightly below the target sample of 900.[12]

The baseline survey was administered immediately before the experimental treatments, and asked questions about participants' demographics, past experience with vote-selling, expectations about monetary offers, and preference ratings for the candidates for mayor, vice-mayor, and city council. We also asked participants to rate each candidate for mayor, vice-mayor, and city council according to how favorable they felt towards each candidate on a 7-point Likert scale (−3 = extremely unfavorable, 0 = neutral, 3 = extremely favorable).

After completing the baseline survey, all participants were shown a three-minute video clip on the hand-held device. The video clip is a humorous appeal to sensible electoral participation, encouraging viewers to turn out to vote, vote for honest and competent candidates, and consider the public good in their voting decisions. In this context, it includes an appeal to voters not to sell their votes.[13] The video clip was shown to all respondents to ensure that those in control and treatment groups received similar appeals not to sell their votes. This is important because the promise treatments, by themselves, might be construed as including an implicit suggestion not to sell one's vote. Our interest is in evaluating the effectiveness of the promise treatments themselves, over and above appeals to eschew vote-selling. Making an explicit appeal to all respondents not to sell their votes (by showing the video) helps sharpen the interpretation of the treatments as being due to the promises elicited, and not due to any appeal not to sell one's vote that might be perceived as bundled with the promise elicitation.

### 3.2. Treatments

At the end of the voter educational video clip, respondents in the two treatment groups (Promise 1 and Promise 2) were invited to make promises not to sell their votes in the upcoming election, in ways that differed across the treatments.[14] Individuals in the Promise 1 treatment were asked to make a promise not to accept money from any candidate, while those in the Promise 2 treatment were asked to promise to vote according to their conscience even if they accepted money.

Elicitation of the promises was implemented by showing respondents a screen on the hand-held device. For Promise 1, the screen is reproduced as Fig. 1a. The text on the screen reads: "Would you promise not to take the money from any candidate or local leader before the elections?" For Promise 2, the screen image can be seen as Fig. 1b, and the corresponding text is: "If any candidate or local leader gives you money before the elections and you decide to keep it, would you promise to vote according to your conscience?"

On both screens, participants were asked to tap on either of the images shown in the figures to register their response. Tapping the left image (of a handshake, above the words "Yes, I promise.") would signify agreement to promise, and tapping the right image (of an open hand in a "halt" signal, above the words "No, I can't make that promise.") would indicate refusal to promise.

A participant who agreed to make the promise by tapping on the image of the handshake was then asked, on the next screen, to write the words "I promise" on a blank space using their finger (see Fig. 1c).[15] After the signature, participants were asked two additional questions on politics and vote-buying, and the survey ended.[16]

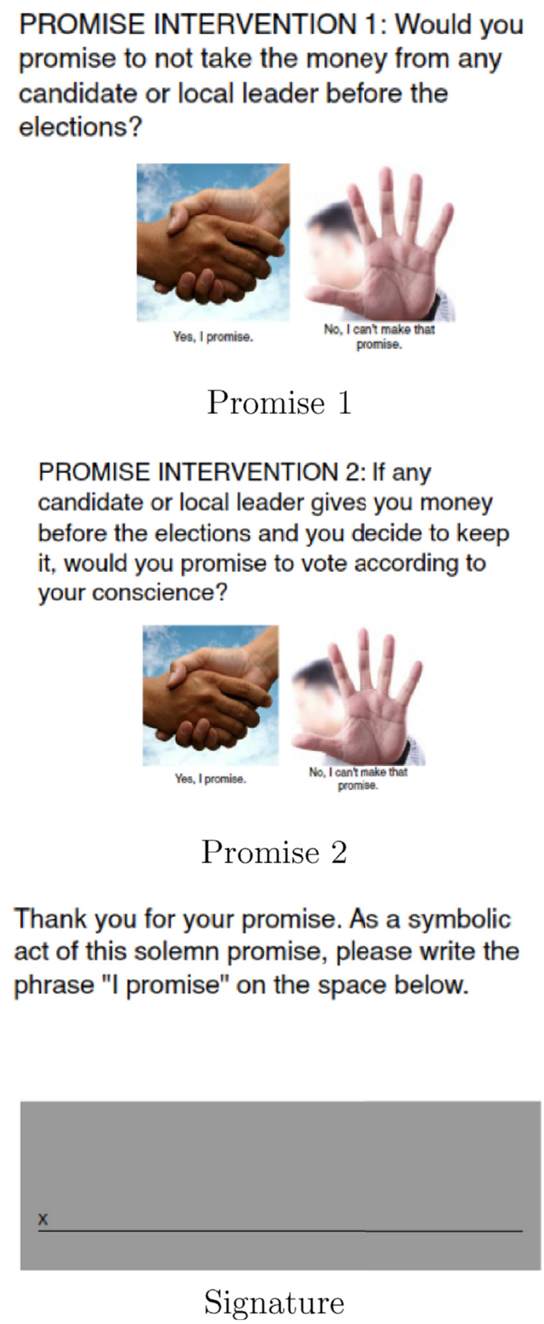



**Fig. 1.** Promise treatments as viewed by participants.

[11] The list of alternates was sorted according to a randomly assigned number. When replacing primary respondents who could not be interviewed, enumerators picked alternates in the prescribed randomized order.

[12] In total, enumerators sought to locate 1496 voters. Reasons for unsuccessful baseline surveys were as follows: failed to contact after repeated visits (170 voters), out of town (154), migrated out of Sorsogon City (92), refused (65), moved to unknown location (65), deceased (21), and other (27). This led to 902 voters being administered the baseline survey. Of these, 19 provided incomplete baseline responses, yielding our baseline sample of 883.

[13] The video clip features Mae Paner, a political activist and actress, as the fictional character "Juana Change." The video can be viewed here: http://www.youtube.com/watch?v=10Jh8Nzu7Zs.

[14] The majority of vote-buying in Sorsogon happens in the few days in advance of the election, so the treatments were administered, on average, two weeks prior to the period when vote-buying payments were made.

[15] On that screen, the text read "Thank you for your promise. As a symbolic act of your solemn promise, please write the phrase 'I promise' on the space below".

[16] These last two questions were also asked of participants who refused to promise as well as of those in the control group (who were not asked to make any promises).

### 3.3. Post-election survey

We fielded an endline survey of the same study participants from May 17 to June 8, 2013, a period spanning 4–26 days after the May 13, 2013 midterm election. The endline survey collected data on whether respondents voted (turnout), as well as which candidates they voted for in each race (mayor, vice-mayor, and city council). We achieved a high (95.9%) endline survey success rate, and this rate is not differential by treatment status (as discussed further below).

### 3.4. Initial hypotheses

Based on previous research on promises and informal agreements, we anticipated that both promises would be effective in reducing vote-selling (Charness and Dufwenberg, 2006; Vanberg, 2008; Kessler and Leider, 2012; Krupka et al., 2013). If voters have made the promise not to sell their vote, we expect that the social norm that it is important to keep one's promises will cause many voters to follow their promise, and either turn down offers of money, or vote for their preferred candidate even if they receive money from other candidates. We expected that the primary difference between treatments would be in the uptake of the promise, with more voters predicted to make Promise 2 — since those voters could still accept money without breaking their promise.

## 4. Empirical strategy

### 4.1. Proxy for vote-selling

The Philippines has a secret ballot, so measurement of vote-selling behavior is a first-order challenge. We did not ask participants directly whether they sold their votes, due to concerns about experimenter demand or social desirability bias. If individuals in the treatment groups underreported the extent to which they sold their votes, this would lead to spurious findings that the treatments reduced vote-selling.

Our approach instead is to simply ask participants in the endline who they voted for in the individual races (for mayor, vice-mayor, and city council), and to compare their reported votes in the endline survey with the candidate favorability ratings they reported in the baseline survey. Our key outcome variable is vote switching: an indicator equal to one for a particular election race if the respondent reported in the endline survey that they voted for a candidate who they did not rank highest in the baseline survey for that position (in the Likert-scale elicitation), and zero if they did say in the endline that they voted for their highest-rated candidate. We construct vote-switching indicators for each race separately, as well as indicators of whether the voter switched in any race.[17]

There are a number of reasons other than vote-selling why a voter may have voted for a candidate other than his or her top-rated candidate (e.g., learning new information.) We expect that such "legitimate" reasons should be unaffected by the promise interventions. Therefore, *differences* in vote-switching across treatment conditions should represent differences in vote-selling.[18]

A few comments are in order regarding the use of vote switching as a proxy for vote-selling. First of all, it is important that candidate favorability ratings in the baseline must provide an unbiased indication of participants' true preferences for candidates. This is likely to be satisfied: our survey staff presented themselves as neutral and unaffiliated with any candidate or political party, and favorability ratings were elicited before respondents were exposed to any of our promise treatments.

Second, it is important that our vote switching measure take into account bias that might be due to social desirability. It might be the case that voters are reluctant to appear to have broken a promise, and so they may be less willing to report voting for candidates that gave them money. In Section 6, we formalize social desirability bias within our model, and discuss empirical tests (comparisons of treatment effects across promises and across races) that are robust to the presence of social desirability bias. However, there is also an *a priori* case to be made that social desirability bias may not be large to begin with. In the endline survey we did not remind respondents that we had data on their candidate ratings from the baseline survey. It also requires a fair degree of sophistication for a respondent to recall that our project was about vote-selling, to recall their candidate preferences from the baseline, and to intentionally misreport their votes to be consistent with their original preferences. What's more, it should be far from clear to respondents that reporting in the endline that one voted for someone who was not their highest-rated choice in the baseline would be viewed by enumerators as ethically questionable, because vote-switching could occur for legitimate reasons (as mentioned above). Respondents, in essence, have "cover" to report at endline that they voted for someone other than their initially-preferred candidate, since such switches can occur for many reasons other than vote-selling.

In other work we analyze the plausibility of the vote switching measure by assessing whether it corresponds to relationships outside of the scope of our theory (Hicken et al., 2015). For example, as we would expect, we find that switching rates are higher when more money is offered, and voters are more likely to switch the narrower the gap in preference between their most preferred candidate and the next best alternative.

If one believes that vote-switching is an acceptable proxy for vote-selling, and if our interventions are effective at reducing vote-selling, respondents in the treatment groups should do less vote switching (as defined above). If, on the other hand, any of our treatments led to more vote-selling, we should see increases in the rate of vote switching.

### 4.2. Regression specification

We assess the effect of the promises on vote-switching by estimating the following ordinary-least-squares regression equation (a linear probability model):

$$y_{ij} = \alpha + \beta_{1j}\, Promise1_i + \beta_{2j}\, Promise2_i + X_i'\gamma + \epsilon_{ij} \tag{1}$$

$y_{ij}$ is an indicator variable equal to 1 if the respondent switched his or her vote in race $j$, and 0 otherwise. $Promise1_i$ and $Promise2_i$ are indicator variables equal to 1 if the respondent was randomized into (respectively) the Promise 1 or Promise 2 treatment, and 0 otherwise. $X_i'$ is a vector of baseline (pre-treatment) control variables. $\epsilon_{ij}$ is a mean-zero error term. We report robust (Huber/White) standard errors.

The coefficients of interest are $\beta_{1j}$ and $\beta_{2j}$ on the treatment indicators, which measure (respectively) the impact of treatment on the

[17] For the city council race, in which each voter casts votes for four candidates, vote switching is defined as voting for at least one candidate who was not among their top four rated candidates. We define vote switching in the city council race in this way (as a dummy variable) to maintain comparability with the mayor and vice-mayor switching variables. In this sense, the city council variable captures the extensive margin of vote switching in that race. One can also examine the intensive margin of vote switching in the city council race. Regression results are robust to defining this latter variable as the number of people the respondent voted for who were not in their list of top four candidates prior to the election (which takes on integer values between 0 and 4 inclusive.).

[18] Another possibility is that voter preferences and actual voting could be misaligned due to strategic voting (Alvarez and Nagler, 2000). However, we find no obvious reason to believe that the promise treatments would affect strategic voting, so we simply consider strategic voting as another determinant of vote-switching that should be orthogonal to our treatments.

probability of vote-switching. To be clear, because making the promise is endogenous, we focus here on the effect of being in the promise treatment (being invited to make a promise), and not on whether the respondent actually made the promise. Our estimates are therefore intent-to-treat (ITT) effects.

## 5. Results

### 5.1. Summary statistics, baseline balance, and promise take-up

Panel A of Table 1 reports summary statistics for key baseline variables, in the full sample (column 1) and in the subsamples by treatment condition (columns 2–4). The columns to the right report, for each baseline variable, the p-values of F-tests of the joint equality of means across treatment conditions as well as for pairwise combinations of the treatment conditions. There is no indication of substantial imbalance in baseline characteristics across treatment conditions. Out of 100 p-values shown in Panel A, 10 are below 0.10, which is exactly the proportion that would be expected to occur by chance. To account for any biases generated by these chance imbalances, these baseline variables will be included as control variables in the regressions.[19]

Panel B of Table 1 reports similar summary statistics for promise-making and the key dependent variables of interest. The first row of this panel reports the fraction of respondents making the elicited promises in each treatment group. In each treatment group, slightly more than half of respondents make the promise—51% for Promise 1 ("Don't take the money") and 56% for Promise 2 ("Take money, vote conscience")—and these proportions are not different from one another at conventional levels of statistical significance.

### 5.2. Vote-shares and candidate favorability ratings

Table 2 provides relevant data for each candidate and electoral race. Candidates in bold are winners of their respective races, and starred candidates are incumbents. Reported vote shares in our sample (from the endline survey) correctly predict the actual winners in each race. The correlation coefficient between actual and sample-reported vote shares (columns 1 and 2) is 0.957.

Average favorability ratings across candidates from our endline survey are also highly correlated with vote shares. The correlation coefficient between the average favorability rating (column 3) and reported vote share in the sample (column 2) is 0.838 (and the corresponding correlation with the actual vote share in column 1 is 0.839). The remaining columns of the table display the distribution of discrete candidate favorability ratings, across the integers ranging from −3 to 3. There is considerable variation in candidate favorability ratings across our survey respondents across the range of possible responses.

### 5.3. Attrition from baseline to endline surveys

To be included in the endline sample for analysis of a particular electoral outcome, a baseline respondent had to have: 1) completed the endline survey, 2) actually turned out to vote in the election, and 3) reported who they voted for in a given electoral race. If either treatment affected attrition (on any of these margins), one might worry that any observed treatment effects on vote-switching could be simply due to compositional changes in the sample. Out of the 883 baseline respondents, the share who completed the endline survey, voted, and reported their mayoral vote was 86.0%. The corresponding shares for vice-mayor and city council are 85.0% and 90.0%.

Differences in these measures of attrition across treatment conditions are very small, and none are statistically significantly different from zero, so attrition bias is of little concern in this context.[20] Please refer to Online Appendix Table 1 for further details.

Voter turnout in particular is of interest, because any treatment effects on turnout (among those surveyed at follow-up) could be a source of selection bias. As shown in column 2 of Online Appendix Table 1, turnout in the control group is 93.8%, while turnout in the treatment groups is not substantially or statistically significantly different: for the Promise 1 treatment, turnout is lower by 1.03 percentage points (standard error 2.06 percentage points), while in the Promise 2 group turnout is 0.04 percentage points lower (standard error 1.96 percentage points). One point of note is that the reported turnout rate in our sample appears high. In comparison, voter turnout in Sorsogon City in the same 2013 election that we study is 79.4%.[21] There are at least two reasons why this might be the case. First of all, our sample probably selected for people who have a lower opportunity cost of time, because they were home at the time our survey enumerators showed up, and agreed to participate in the survey. Such individuals, because of their lower time constraints, may also have higher voter turnout in general. Second, it is possible that inclusion in our study may have led to higher voter turnout, perhaps because of the voter-education video we showed to all respondents and/or because the questions we asked in the survey raised their desire to vote in the coming election. Because turnout is balanced across treatment conditions, these factors do not threaten the internal validity of the study, but do suggest that the survey sample itself may be composed somewhat differently than the general voting population.

### 5.4. Impact of treatments on vote-switching

We first present our results in graphical form. Fig. 2 displays the bar graphs of the percentage of vote switching, by treatment condition, with 95% confidence intervals. Fig. 2a presents the share of respondents who switched votes in at least one of the races. In the control group, 57.4% of subjects switched their vote at least once, compared to 50.4% in the Promise 1 treatment, and 61.8% in the Promise 2 treatment. This provides a first indication that the promise treatments had opposite effects, with asking subjects not take money from candidates reducing the amount of vote switching, while asking subjects to vote their conscience even if they take money increases the amount of vote switching.

Fig. 2b and c examine vote switching separately in, respectively, the mayor/vice-mayor races and the city council race. In the mayor/vice-mayor races, vote switching rates are very similar in the control and Promise 1 groups (26.4% and 27.1%, respectively), but higher in the Promise 2 group (33.7%). By contrast, for the city council race, the control and Promise 2 groups have similar vote switching rates (47.1% and 47.7%, respectively), while the rate for the Promise 1 group is much lower, at 38.4%.

To confirm these visual impressions, we now turn to estimation of regression Equation (1) for vote switching in different races. Results are presented in Table 3.

In column 1, the dependent variable is vote switching in any race. As in Fig. 2a, the coefficient for the Promise 1 treatment is negative, while the coefficient for Promise 2 is positive. The negative coefficient on Promise 1 is statistically significantly different from zero at the 5% level.

[19] Our results are robust to exclusion of the baseline control variables. Regression results and tests of theoretical predictions when control variables are not included in the regressions are presented in Online Appendix Tables 2 and 3, and should be compared with Tables 3 and 4 of the main text.

[20] In results available upon request, we also find that, among respondents completing the end line survey, there is no effect of either promise treatment on turning out to vote.

[21] Based on the official statement of votes for Sorsogon City that we obtained from the Commission on Elections.

**Table 1**
Baseline survey summary statistics and balance tests.

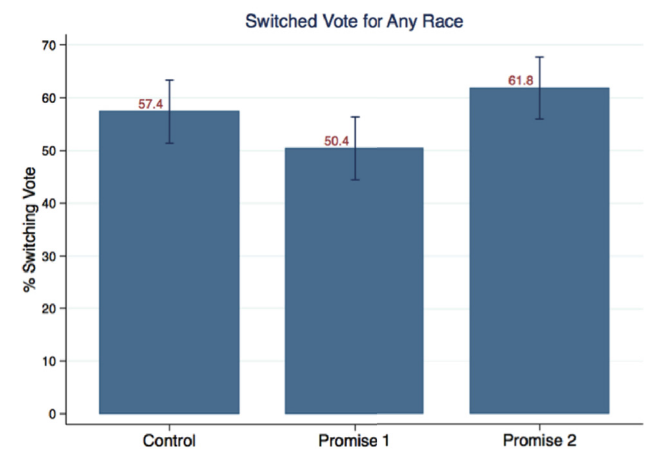
| | Treatment groups | | | | *p–values* | | | |
|---|---|---|---|---|---|---|---|---|
| | Full sample | Control group | Promise 1 ("Don't take money") | Promise 2 ("Take money, vote conscience") | $C = P1 = P2$ | $C = P1$ | $C = P2$ | $P1 = P2$ |
| Number of observations | 883 | 291 | 298 | 294 | | | | |
| **Panel A: Baseline variables** | | | | | | | | |
| Male (indicator) | 0.450 | 0.471 | 0.426 | 0.452 | 0.550 | 0.277 | 0.656 | 0.521 |
| Years of age | 42.02 (16.29) | 43.56 (17.15) | 41.61 (16.35) | 40.90 (15.25) | 0.132 | 0.159 | 0.049 | 0.587 |
| Religion is Catholic (indicator) | 0.922 | 0.911 | 0.926 | 0.929 | 0.697 | 0.492 | 0.426 | 0.911 |
| Number of voting household members | 3.55 (1.93) | 3.62 (2.14) | 3.62 (1.81) | 3.42 (1.84) | 0.319 | 0.994 | 0.218 | 0.177 |
| Single | 0.258 | 0.251 | 0.269 | 0.255 | 0.879 | 0.627 | 0.906 | 0.712 |
| Married | 0.526 | 0.526 | 0.517 | 0.534 | 0.916 | 0.827 | 0.842 | 0.675 |
| Widowed | 0.075 | 0.083 | 0.071 | 0.071 | 0.836 | 0.584 | 0.617 | 0.964 |
| Domestic partnership | 0.123 | 0.117 | 0.138 | 0.116 | 0.673 | 0.451 | 0.964 | 0.423 |
| Separated | 0.018 | 0.024 | 0.007 | 0.024 | 0.094 | 0.088 | 0.985 | 0.091 |
| Choose not to work | 0.239 | 0.227 | 0.285 | 0.204 | 0.064 | 0.104 | 0.505 | 0.022 |
| Retired | 0.046 | 0.065 | 0.030 | 0.044 | 0.135 | 0.046 | 0.263 | 0.369 |
| Student | 0.045 | 0.048 | 0.044 | 0.044 | 0.962 | 0.795 | 0.823 | 0.972 |
| Unemployed, looking | 0.099 | 0.089 | 0.107 | 0.099 | 0.763 | 0.463 | 0.701 | 0.727 |
| Working full-time | 0.324 | 0.357 | 0.269 | 0.347 | 0.035 | 0.020 | 0.792 | 0.039 |
| Working part-time | 0.247 | 0.213 | 0.265 | 0.262 | 0.246 | 0.139 | 0.165 | 0.930 |
| Some elementary to no schooling | 0.120 | 0.127 | 0.114 | 0.119 | 0.888 | 0.627 | 0.766 | 0.756 |
| Elementary | 0.176 | 0.151 | 0.201 | 0.174 | 0.279 | 0.110 | 0.466 | 0.386 |
| Some highschool | 0.193 | 0.210 | 0.198 | 0.170 | 0.448 | 0.727 | 0.223 | 0.381 |
| Highschool | 0.168 | 0.165 | 0.161 | 0.177 | 0.869 | 0.899 | 0.702 | 0.609 |
| Some college | 0.131 | 0.131 | 0.111 | 0.153 | 0.314 | 0.461 | 0.437 | 0.129 |
| College up | 0.174 | 0.189 | 0.171 | 0.163 | 0.438 | 0.291 | 0.277 | 0.972 |
| Vocational | 0.039 | 0.028 | 0.044 | 0.044 | 0.708 | 0.573 | 0.415 | 0.798 |
| Born here | 0.727 | 0.715 | 0.745 | 0.721 | 0.682 | 0.410 | 0.866 | 0.512 |
| Migrated as a child | 0.107 | 0.107 | 0.104 | 0.109 | 0.982 | 0.921 | 0.928 | 0.850 |
| Migrated as an adult | 0.167 | 0.179 | 0.151 | 0.170 | 0.646 | 0.366 | 0.784 | 0.528 |
| **Panel B: Promise-making outcome variables** | | | | | | | | |
| Made promise (indicator) | – | – | 0.514 | 0.557 | – | – | – | 0.295 |
| Switched Vote in Any Race (indicator) | 0.565 | 0.574 | 0.504 | 0.618 | 0.026 | 0.104 | 0.298 | 0.007 |
| Switched Vote for Mayor (indicator) | 0.123 | 0.106 | 0.115 | 0.146 | 0.373 | 0.729 | 0.172 | 0.301 |
| Switched Vote for Vice-Mayor (indicator) | 0.220 | 0.206 | 0.198 | 0.256 | 0.250 | 0.823 | 0.183 | 0.118 |
| Switched Vote for City Council (indicator) | 0.444 | 0.471 | 0.384 | 0.477 | 0.052 | 0.043 | 0.891 | 0.030 |

Notes: Values in the first four columns are means (standard deviations). Variables in Panel A collected in baseline survey, administered from April 17 to May 8, 2013 (prior to May 13, 2013 municipal elections). Promises (first variable in Panel B) were elicited at end of baseline survey. Remaining variables in Panel B are dependent variables in the analysis, and were constructed on the basis of reported voting in endline survey (May 17 to June 8, 2013). Respondents randomized with equal (1/3) probability into the control group, Promise 1 treatment group, or Promise 2 treatment group. P-values are for F-tests that mean of variable is equal across the specified treatment conditions.

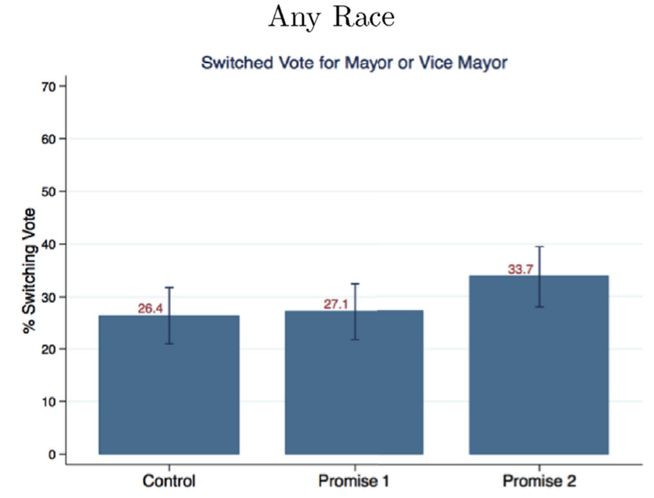
**Table 2**
Vote shares and candidate favorability ratings, by electoral race.

| Candidate | Actual vote share in election | Reported vote share (endline survey) | Sample average favorability rating | % of surveyed respondents rating candidates as… | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Extremely unfavorable (−3) | Quite unfavorable (−2) | Slightly unfavorable (−1) | Neutral (0) | Slightly favorable (1) | Quite favorable (2) | Extremely favorable (3) |
| Mayor race | | | | | | | | | | |
| A* | 46.3 | 44.8 | 0.5 | 5.9 | 10.7 | 7.4 | 29.2 | 13.9 | 21.3 | 11.7 |
| **B** | 48 | 55.2 | 0.6 | 3.2 | 8 | 6.5 | 30.7 | 21.7 | 21.1 | 8.8 |
| Vice-mayor race | | | | | | | | | | |
| C | 30 | 30.2 | 0.3 | 2.5 | 17 | 7.8 | 27.5 | 19.6 | 20.3 | 5.3 |
| D | 24.6 | 23.4 | 0 | 3.7 | 19.8 | 8.7 | 33.6 | 16.1 | 13.6 | 4.4 |
| **E** | 32.9 | 46.5 | 0.4 | 3.1 | 13.7 | 6.3 | 29.8 | 18.2 | 21.9 | 7 |
| City council race, Bacon District | | | | | | | | | | |
| F | 30.4 | 31 | 0.1 | 2.1 | 21.6 | 7 | 29.6 | 18.5 | 17.8 | 3.5 |
| G | 24.5 | 24.5 | −0.2 | 4.9 | 25.1 | 7 | 33.1 | 12.5 | 13.9 | 3.5 |
| **H*** | 37.8 | 46 | 1 | 1.1 | 8.4 | 2.4 | 24 | 19.5 | 35.5 | 9.1 |
| I | 10.2 | 11.1 | −0.3 | 3.8 | 27.9 | 8 | 35.2 | 8.7 | 14.3 | 2.1 |
| **J** | 32.7 | 44.4 | 0.6 | 1.1 | 13.2 | 5.9 | 30.7 | 15.3 | 25.1 | 8.7 |
| **K** | 32.5 | 39.9 | 0.4 | 2.4 | 15.7 | 4.9 | 27.5 | 19.9 | 22 | 7.7 |
| **L*** | 37.3 | 54 | 0.7 | 1.7 | 13.2 | 7 | 21.6 | 16.4 | 32.1 | 8 |
| M | 15.8 | 17.6 | −0.1 | 2.8 | 23.3 | 8.4 | 32.4 | 15 | 16 | 2.1 |
| N | 17.1 | 18.8 | 0 | 2.8 | 23.3 | 5.6 | 30.7 | 15.3 | 17.4 | 4.9 |
| O* | 25.5 | 24.9 | 0.2 | 2.4 | 21.3 | 5.2 | 30.3 | 15.3 | 16.7 | 8.7 |
| P | 20.1 | 14.6 | −0.2 | 4.2 | 26.8 | 6.6 | 32.4 | 12.2 | 14.6 | 3.1 |
| City council race, East District | | | | | | | | | | |
| Q* | 31.3 | 28.7 | 0.3 | 1.8 | 18.7 | 7.8 | 26.5 | 15.9 | 26.5 | 2.8 |
| R* | 20.4 | 23 | 0.6 | 1.1 | 13.8 | 6 | 24 | 21.2 | 28.3 | 5.7 |
| S | 6.6 | 4.6 | −0.6 | 4.6 | 33.9 | 8.5 | 35 | 10.3 | 7.1 | 0.7 |
| T* | 29.7 | 33 | 0.8 | 0.4 | 12 | 3.2 | 23 | 23.3 | 31.1 | 7.1 |
| U | 3.2 | 1.5 | −0.8 | 2.8 | 36.8 | 11 | 35.7 | 10.6 | 3.2 | 0 |
| V | 5.5 | 1.2 | −0.6 | 2.1 | 32.5 | 11.3 | 39.9 | 10.3 | 3.5 | 0.4 |
| **W** | 40.5 | 60.2 | 0.6 | 0.4 | 13.1 | 4.6 | 25.8 | 21.2 | 30 | 5 |
| **X** | 45.4 | 60.2 | 0.8 | 0.7 | 12 | 6.4 | 24.4 | 14.8 | 28.6 | 13.1 |
| **Y** | 44.4 | 54 | 0.3 | 0.4 | 20.1 | 6.4 | 29.3 | 15.9 | 23.7 | 4.2 |
| **Z** | 34.5 | 39.5 | 0.1 | 1.4 | 21.6 | 9.9 | 27.9 | 19.1 | 16.3 | 3.9 |
| AA | 15.1 | 13.8 | 0 | 0.7 | 20.9 | 7.8 | 35.7 | 17.7 | 14.8 | 2.5 |
| AB | 9.2 | 4.2 | −0.5 | 2.5 | 31.1 | 9.5 | 35.7 | 12.4 | 8.5 | 0.4 |
| AC | 18.2 | 11.9 | 0.5 | 1.4 | 17.7 | 2.8 | 24.7 | 22.3 | 25.4 | 5.7 |
| City council race, West District | | | | | | | | | | |
| AD | 5 | 4.2 | −0.1 | 2.6 | 19.8 | 11.2 | 37.7 | 11.8 | 14.7 | 2.2 |
| AE | 20.1 | 13.4 | 0.2 | 2.9 | 14.4 | 7.4 | 35.5 | 19.8 | 18.5 | 1.6 |
| **AF*** | 49.8 | 60.8 | 1.1 | 0 | 8 | 1.9 | 26.8 | 14.7 | 36.4 | 12.1 |
| AG | 32.7 | 43.8 | 0.4 | 0.3 | 13.4 | 7.7 | 34.8 | 16.9 | 24 | 2.9 |
| AH | 18.2 | 19.1 | 0.3 | 1.6 | 17.3 | 5.4 | 38.3 | 11.8 | 17.3 | 8.3 |
| **AI** | 37.3 | 50.9 | 0.9 | 1.9 | 10.9 | 2.9 | 23.6 | 12.8 | 36.4 | 11.5 |
| **AJ*** | 38.4 | 59 | 0.7 | 1.3 | 10.2 | 4.8 | 29.1 | 19.2 | 29.7 | 5.8 |
| AK | 27.4 | 23.7 | 0.5 | 1 | 12.1 | 5.1 | 34.2 | 17.3 | 21.1 | 9.3 |
| **AL*** | 34.6 | 50.9 | 0.7 | 1 | 9.6 | 4.5 | 29.4 | 18.2 | 31.6 | 5.8 |
| AM | 3.1 | 4.2 | −0.5 | 1.9 | 28.1 | 11.2 | 43.5 | 9 | 5.8 | 0.6 |
| AN | 9.7 | 7.4 | −0.3 | 1.9 | 23.6 | 9.6 | 43.1 | 13.1 | 7.4 | 1.3 |
| AO | 8.3 | 4.6 | −0.3 | 2.9 | 24.3 | 8.3 | 45.7 | 7 | 9.6 | 2.2 |

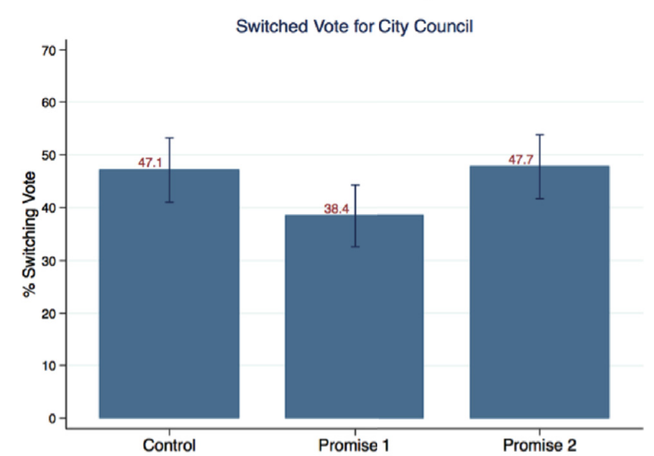
Notes: Data on actual vote share in election are from Philippine Commission on Elections (COMELEC). Reported vote share is from our endline survey. Favorability ratings are from our baseline survey. Starred (*) candidates are incumbents (but not all incumbents ran again in this election). Bold candidates are winners of their respective races. In city council races, top four candidates are elected.




Any Race




Mayor or Vice-Mayor




City Council

**Fig. 2.** Vote-Switching by Treatment Condition. Notes: Figures show fraction of respondents switching their vote (voting for a candidate other than their top-rated candidate as reported in baseline survey), by treatment condition, along with % confidence intervals. **(a)** shows fraction vote switching in any of the three races. **(b)** shows fraction vote switching in either of the mayor or vice-mayor races. **(c)** shows fraction vote switching in the city council race. (In city council race, voters can vote for up to four candidates. Vote switching in this race is defined as voting for at least one city council candidate who was not among the respondent?s top four rated candidates in baseline survey.)

In columns 2–4 we examine treatment effects in specific races (mayor, vice-mayor, and city council, respectively). These results reveal that the estimated effect on overall vote switching estimated in column 1 obscures heterogeneity of treatment effects across races.

In regressions for vote-switching in the mayor or vice-mayor races, the coefficient on Promise 1 is always relatively small in magnitude and negative in sign. The coefficient on Promise 2, on the other hand, is larger in magnitude, and positive in sign in both cases. None of these promise treatment effects for the mayor or vice-mayor races are statistically significantly different from zero at conventional levels.

The pattern is quite different in analysis of vote switching in the city council race (column 4). The Promise 1 treatment has a large, negative and statistically significant (at the 5% level) effect on vote switching, amounting to a 10.9 percentage point reduction. The corresponding coefficient for Promise 2 is also negative but, by contrast, is very small in magnitude and is not statistically significantly different from zero.

### 5.5. *Some thoughts on social desirability bias*

Examination of individual treatment coefficients may be misleading, because social desirability bias may lead estimated promise treatment effects to be biased in a negative direction (i.e., for treatments to appear to reduce vote switching). In general, studies have found that social desirability bias regarding vote selling/vote buying is low in the Philippines (Hicken et al., Forthcoming). Both voters and candidates freely admit to participating in the practice and there is little social stigma attached to such a confession (Hicken et al., Forthcoming). In fact, survey list experiments designed to estimate the magnitude of social desirability bias show that it is negligible (Cruz et al., 2015). Nonetheless, in Section 6 below we formalize what effect social desirability should have, if it exists, and present tests of differences in treatment effects across promises and across races that are robust to the presence of social desirability bias.

## 6. Behavioral explanations

### 6.1. *Vote-selling as temptation*

We present here one potential behavioral mechanism that is consistent with the results we observe: that voting for a candidate that gave money represents a kind of temptation, and that the promises can change voters willingness to expose themselves to the temptation. We describe the intuition behind a simple model of vote-selling and the impact of promises not to sell one's vote, with the formal details of the model presented in the Online Appendix. We discuss which features of the model are important to explaining our results, and examine other potential models that cannot generate our results.

We take as a starting point the findings of Finan and Schechter (2012) that vote-buying operates through a reciprocity channel. This is an appropriate assumption in our setting, since as described previously the Philippines uses electronic balloting and therefore candidates cannot verify the vote of an individual voter. However, we consider a voter's inclination to reciprocate a candidate's gift through voting as a temptation present in the moment, rather than an intrinsic permanent part of the voter's utility. If the reciprocity of vote-selling was a preference of the voter, then the promises we study can only have a beneficial effect, and the promise to "take money but vote your conscience" (Promise 2) would dominate the promise not to accept money (Promise 1), both in terms of uptake and in effect on voting.

By contrast, if vote-selling is a temptation problem present in the voting booth, then a (sophisticated) voter will think carefully about whether to accept a candidate's offer. Specifically, when a voter is offered a gift, he can (partially) anticipate how that gift will affect his future vote. The voter wants to take money, and wants to vote for his ex-ante preferred candidate, but does not intrinsically value the reciprocating the gift. Therefore, the voter will happily accept gifts from his

**Table 3**
Impact of treatments on vote-switching (ordinary least-squares regressions).

| Dependent variable: | | Switched Vote in Any Race (1) | | Switched Vote for Mayor (2) | | Switched Vote for Vice-Mayor (3) | | Switched Vote for City Council (4) |
|---|---|---|---|---|---|---|---|---|
| Promise 1 treatment ("Don't take money") | $\beta_1$ | −0.0953** (0.0429) | $\beta_{1m}$ | −0.00329 (0.0278) | $\beta_{1v}$ | −0.0221 (0.0365) | $\beta_{1c}$ | −0.109** (0.0430) |
| Promise 2 treatment ("Take money, vote conscience") | $\beta_2$ | 0.0309 (0.0427) | $\beta_{2m}$ | 0.0288 (0.0299) | $\beta_{2v}$ | 0.0391 (0.0383) | $\beta_{2c}$ | −0.00945 (0.0439) |
| Control variables | | Y | | Y | | Y | | Y |
| Mean of dependent variable (control group) | | 0.5736 | | 0.1057 | | 0.2056 | | 0.4713 |
| Observations | | 806 | | 759 | | 751 | | 793 |
| R-squared | | 0.046 | | 0.037 | | 0.041 | | 0.042 |

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
Notes: Robust (Huber/White) standard errors in parentheses. Dependent variable in columns 1–10 equal to 1 if respondent switched his/her vote in the given race or set of races, 0 otherwise. Vote switching in mayor and vice-mayor races defined as voting for a candidate not receiving respondent's highest favorability rating in baseline (pre-election) survey. Vote switching in city council race defined as voting for a candidate not among the respondent's top-four highest-favored candidates in baseline survey. Respondents randomized with equal (1/3) probability into the control group, Promise 1 treatment group, or Promise 2 treatment group. Control variables are listed in Panel A of Table 1 and were reported in baseline survey prior to treatment.

preferred candidate, gifts from another candidate that are too small to change his vote, or gifts that are so large that it outweighs the cost of the vote switch. However, a voter may turn down an intermediate gift that is large enough to change his vote in the moment, but too small to be worth that cost. In modeling temptation we consider three cases: fully sophisticated (the voter correctly anticipates his future temptation), fully naïve (the voter believes he will not face temptation), and partially sophisticated (the voter anticipates future temptation, but underestimates its magnitude).

What role does a promise play? We assume that having made a promise, voters receive utility from actions consistent with the promise, and disutility from violating the promise. For Promise 1, if the voter makes the promise it will cause him to turn down gifts that aren't too large. This reduces the amount of temptation he faces in the voting booth, and therefore reduces the overall level of vote switching. Additionally, if the utility of making and following the promise is large enough (relative to the offered gifts expected), then the voter is willing to make the promise.

Promise 2 can similarly reduce vote selling if the utility from keeping a promise is large enough. This can occur both from more voters turning down gifts that would lead to a vote change, and from the promise being strong enough to keep voters from switching even after they accept the gift. However, for partially sophisticated voters, Promise 2 can also *increase* the amount of vote switching. This can happen if the voter is sophisticated enough to turn down the gift from a less-preferred candidate absent a promise, but naïve enough to mistakenly believe that the promise will prevent him from switching. This occurs for intermediate gift sizes, with more voters accepting gifts than in the base case, and then being surprised by the strength of the temptation and ultimately switching their votes. Furthermore, this negative effect for the promise to vote your conscience can occur simultaneously with more positive effects for other promises and races.

### 6.2. Predictions of the temptation model

Based on this model, we can identify several predictions comparing vote-switching behavior between promises and races. Recall that the empirical analysis (reported in Table 3) estimated the impact of each promise on vote-switching in three different electoral races, with $\beta_{ij}$ representing the impact of promise $i \in [1,2]$ on vote-switching in electoral race $j \in [m, v, c]$ (mayor, vice-mayor, and city council). To further address the potential impact of social desirability bias (discussed below), we also emphasize predictions focusing on the *difference* between impacts of Promises 1 and 2 ($\beta_{1j} - \beta_{2j}$).

#### 6.2.1. Comparing between treatments

We first identify the predicted treatment effect of each promise for fully sophisticated, fully naïve and partially sophisticated voters. Both the fully sophisticated and fully naïve cases make the predictions that (1) both promise treatments should reduce vote-switching ($\beta_{ij} < 0, \forall i, j$), and (2) the impact of the Promise 2 treatment will be larger in magnitude ($\beta_{1j} - \beta_{2j} > 0, \forall j$). Fully sophisticated voters will only make the promise if they will keep it, and (if there is enough heterogeneity in candidate preferences) the additional utility from keeping the promise will matter for some set of voters. For naïve voters the promises either help them avoid temptation they ignored (Promise 1), or helps them overcome the surprise temptation (Promise 2). Promise 2 is predicted to be more effective, since voters can still accept money from their ex ante preferred candidate.

For partially sophisticated voters Promise 1 should reduce vote-switching ($\beta_{1j} < 0, \forall j$). However, as discussed above, Promise 2 can either reduce vote-selling ($\beta_{2j} < 0, \forall j$) or increase it ($\beta_{2j} > 0, \forall j$). The difference in the impact of Promise 1 from Promise 2, $\beta_{1j} - \beta_{2j}$, can therefore be positive or negative. A finding that Promise 1 reduces vote-switching more than Promise 2 does ($\beta_{1j} - \beta_{2j} < 0$) can only be generated by the partially sophisticated case, not the fully sophisticated or fully naïve cases.

The following table summarizes the model's predictions in each case:

| | Partially sophisticated | Fully sophisticated | Fully naïve |
|---|---|---|---|
| $\beta_{1j}$ | $< 0$ | $< 0$ | $< 0$ |
| $\beta_{2j}$ | $> 0$ or $< 0$ | $< 0$ | $< 0$ |
| $\beta_{1j} - \beta_{2j}$ | $> 0$ or $< 0$ | $> 0$ | $> 0$ |

#### 6.2.2. Comparing between races

The model also makes predictions regarding the relative effects of the promises across electoral races that involve different sizes of vote-buying payments (gifts). In our context, the mayor and vice-mayor races involve larger vote-buying payments, compared to the city council races.

For Promise 1 we expect that the treatment will have more negative effects on vote-switching for races that involve smaller vote-buying payments (in other words, the city council race, compared to either the mayor or vice-mayor race): $\beta_{1c} - \beta_{1m} < 0$ and $\beta_{1c} - \beta_{1v} < 0$.

Promise 2 can either have a positive or negative effect on vote-switching, and there is no unambiguous prediction as to the relationship between the Promise 2 treatment effect magnitude and the size of vote-buying payments. The fully naïve, fully sophisticated and partially

**Table 4**
Tests of theoretical predictions.

| | Races pooled | | Mayor race | | Vice-mayor race | | City council race | |
|---|---|---|---|---|---|---|---|---|
| A. Testing predictions of partially sophisticated theoretical case (within race, effects more negative for Promise 1 than Promise 2) | | | | | | | | |
| | $\beta_1 - \beta_2$ | −0.126** | $\beta_{1m} - \beta_{2m}$ | −0.032 | $\beta_{1v} - \beta_{2v}$ | −0.061 | $\beta_{1c} - \beta_{2c}$ | −0.100** |
| | | (0.043) | | (0.030) | | (0.037) | | (0.043) |
| P-value of F-test: $(\beta_{1m} - \beta_{2m} = 0)$ & $(\beta_{1v} - \beta_{2v} = 0)$ & $(\beta_{1c} - \beta_{2c} = 0)$ | | | | | | | | 0.005 |
| B. Testing prediction of differential effects across races (within promise, effects more negative for city council than in either mayor or vice-mayor races) | | | | | | | | |
| Comparing across races, for Promise 1: | | | $\beta_{1c} - \beta_{1m}$ | −0.106** | $\beta_{1c} - \beta_{1v}$ | −0.087* | | |
| | | | | (0.049) | | (0.053) | | |
| Comparing across races, for Promise 2: | | | $\beta_{2c} - \beta_{2m}$ | −0.038 | $\beta_{2c} - \beta_{2v}$ | −0.049 | | |
| | | | | (0.051) | | (0.055) | | |
| P-value of F-test: $(\beta_{1c} - \beta_{1m} = 0)$ & $(\beta_{1c} - \beta_{1v} = 0)$ & $(\beta_{2c} - \beta_{2m} = 0)$ & $(\beta_{2c} - \beta_{2v} = 0)$ | | | | | | | | 0.086 |
| C. All theoretical predictions in A. and B. combined | | | | | | | | |
| P-value of F-test: $(\beta_{1m} - \beta_{2m} = 0)$ & $(\beta_{1v} - \beta_{2v} = 0)$ & $(\beta_{1c} - \beta_{2c} = 0)$ | | | | | | | | |
| $(\beta_{1c} - \beta_{1m} = 0)$ & $(\beta_{1c} - \beta_{1v} = 0)$ & $(\beta_{2c} - \beta_{2m} = 0)$ & $(\beta_{2c} - \beta_{2v} = 0)$ | | | | | | | | 0.008 |

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
Notes: Table reports tests of linear combinations of coefficients suggested by theory. Robust (Huber/White) standard errors in parentheses. $\beta_{ij}$ is impact of promise $i$ on vote-switching in race $j$ in regressions reported in Table 3.

sophisticated cases are all potentially consistent with finding a more negative effect of Promise 2 in the city council election. However, if the Promise 2 treatment leads to an *increase* in vote-switching in races with larger vote-buying payments (the mayor and vice-mayor races) and either no effect or a decrease in vote-switching in the city council race, then this pattern is informative because it only occurs in the partial sophistication case of the model (not the fully sophisticated or fully naïve cases).

### 6.3. Test of theoretical predictions

Based on the theoretical predictions above, Table 4 reports the results of several pairwise coefficient comparisons using the regression results reported in Table 3.

#### 6.3.1. Effects more negative for promise 1 than promise 2 treatment

In Part A of Table 4, we test the prediction of the partially sophisticated case that Promise 1 has a more negative impact on vote-switching (reduces vote-switching more) than does Promise 2. We first conduct this test across treatment effects in the vote-switching regression pooled across races (coefficients in column 1 of Table 3). The difference in coefficients is negative and statistically significant at the 5% level. When conducting this test separately for each race, we find that for each race Promise 1 has a more negative impact than Promise 2: $\beta_{1j} - \beta_{2j} < 0$ in each race $j$. For the city council race, the difference is statistically significantly different from zero at the 5% level.

To test whether the theoretical prediction that Promise 1's impact is more negative than Promise 2's holds across all races considered simultaneously, we conduct an F-test of the joint hypothesis that $\beta_{1m} - \beta_{2m} = 0$ and $\beta_{1v} - \beta_{2v} = 0$ and $\beta_{1c} - \beta_{2c} = 0$. We reject this hypothesis at the 1% level (the p-value, reported in the bottom row of Part A of Table 4, is 0.005). This result provides statistical confirmation that the full set of empirical results is consistent with the partially sophisticated case, and not the fully sophisticated or fully naïve cases of the model.

#### 6.3.2. Effects more negative for city council than in either mayor or vice-mayor races

In Part B of Table 4, we test the prediction that the effect of the Promise 1 treatment on vote-switching will be more negative for the race with the lower vote-buying payments (the city council race) than for the races with higher vote-buying payments (the mayor and vice-mayor races). The prediction regarding differentials in Promise 2's effects across races is ambiguous; Promise 2's effect could be either higher or lower in the city council race compared to the other races.

We conduct pairwise tests of the differential effects of the treatments across electoral races in Part B of Table 4. The results reveal that, within each promise treatment, pairwise differences in treatment effects between the city council regression, on the one hand, and either the mayor or vice-mayor regression, on the other, are all negative in sign. As discussed previously, these differences are statistically significant at the 10% level or better for the Promise 1 comparisons. While the Promise 2 cross-race tests are not statistically significantly different from zero, the negative point estimates for the differences and the positive point estimates on the Promise 2 treatment coefficient (Table 3, columns 2 and 3) can only occur in the partial sophistication case of the model (not the fully sophisticated or fully naïve cases.)

As an overall test whether the prediction that each promise treatment is more negative for the city council race than in the other races, we conduct an F-test of the joint hypothesis that $\beta_{1c} - \beta_{1m} = 0$ and $\beta_{1c} - \beta_{1v} = 0$ and $\beta_{2c} - \beta_{2m} = 0$ and $\beta_{2c} - \beta_{2v} = 0$. This hypothesis is rejected at the 10% level (the p-value, reported in the bottom row of Part B of Table 4, is 0.086).

#### 6.3.3. Test of joint significance of all pairwise treatment effect differences

Finally, we conduct an F-test of the joint significance of all the pairwise tests examined in Parts A and B. We reject at the 1% level the hypothesis that the pairwise treatment effect differences examined in Parts A and B are jointly zero (the p-value is 0.008, reported in Part C of Table 4).

### 6.4. The role of social desirability bias

In our experiment we only observe subjects' self-reports of their votes. As discussed in Section 5.5 one might imagine that subjects may distort their reported votes due to social desirability bias. This would raise the concern that the estimated effects of the promise treatments would be biased in a negative direction (i.e., for the treatments to appear to reduce vote switching). In Section 5.5 we presented some initial evidence for why social desirability bias is not a major concern. In this section we build on the theoretical model we described and evaluated above to explore how such biases could change reported voting, and whether our pattern of results is consistent with expectations of bias. We present here the basic intuition, and give a more formal treatment in the Online Appendix.

First, subjects' initial candidate ratings should be an accurate reflection of their underlying preferences. The candidate ratings occur before subjects know they will be asked to promise not to sell their vote. Hence voters cannot rate a candidate as their favorite in advance so that they can appear to keep their promises by appearing not to switch their votes. Social desirability bias is more likely to appear in the reported vote, affecting our estimate of the rate of vote switching. How might social desirability bias change these reports?

One natural form of social desirability bias would be for any voter to be reluctant to report switching their vote, due to a general norm against vote selling. For example, suppose that only a fraction $p < 1$ of voters who switched accurately report this, with the remainder claiming to have voted for their initially preferred candidate. In this case all of the observed switching rates would be biased towards zero. However, there will only be a difference in observed switching rates between treatment and control if there is a difference in true switching rates. Therefore this form of bias cannot create the treatment differences we observed.

If instead the bias affects only voters asked to make a promise, then the observed switching rates in both promise treatments might be affected. However, comparisons between races can demonstrate that the promise must be having an effect independent of any bias. If the bias affects both races equally, then any difference between races in a promise treatment can only be generated by differences in true switching rates. If voters exhibit different biases depending on the race, we would generally expect larger biases in the more important races than in the less important races.[22] In this case we would expect larger biases for the mayoral and vice-mayoral election compared to the city council election, and hence the apparent effect of the promise would be most beneficial in the higher stakes elections. In order to observe that the promise was more effective in the city council race, voters would have to feel more uncomfortable appearing to break their promise in the city council election than in the mayoral election, which seems unlikely.[23] Furthermore, as before this kind of bias can only make the promises look beneficial, and would never make them look harmful.

If the magnitude of the social desirability bias is likely to be at least as large in the mayoral and vice-mayoral races as in the city council races, then we can use the observed treatment effect in those races to provide an upper bound estimate on the magnitude of the social desirability bias. Suppose that we assume that the true Promise 1 treatment effects are zero for the higher-money races (mayor and vice-mayor) - then the observed difference in vote switching would be the worst case estimate of the magnitude of social desirability bias.[24]

The pattern of Promise 1 treatment effects across races reported in Table 3 indeed suggests that social desirability bias is not a significant concern in our setting. Promise 1 treatment effects in the mayor and vice-mayor races, while both negative, are quite close to zero, and neither are statistically significantly different from zero. The coefficient in the mayor's race is worth particular attention, since the mayor's race is the most important race of the three. The coefficient (−0.003) indicates a reduction in vote-switching of three-tenths of a percentage point, which is very small relative to the control group mayoral election vote-switching rate of 10.6 percent. Assuming no "true" Promise 1 treatment effect on vote-selling in the mayor's race, taking this coefficient seriously would imply p = 0.97 (social desirability bias leads the self-reported vote-switching rate to be 97% of the true vote-switching rate). Given this, we conclude that our estimated treatment effects (the $\beta_{ij}$ estimates) are likely to be minimally affected by reporting bias.[25]

### 6.4.1. *Social desirability bias for turnout*

Another potential form of social desirability bias is a reluctance to report not voting in an election. Norms for voting might exist naturally, and/or they may have been heightened by our video intervention. Such a bias could potentially cause problems interpreting our switching results for Promise 1 (but cannot lead to the directional increase in switching for Promise 2). The potential issue could arise as follows: suppose that our promise intervention is effective in causing participants not to accept money from candidates, but that the effect is primarily to reduce turnout. However, due to the bias against not voting participants report voting for their initially favored candidate. This could then appear as a reduction in our switching measure.

First, note that by this explanation the promise does have an effect in reducing the acceptance of money from candidates. While it is ambiguous which candidates would benefit most from lower turnout races, and reducing turnout is not an ideal outcome, a lower likelihood of voters accepting money should make vote buying less attractive for candidates. Second, in the city council race (where we see the largest effect of the promise), one might expect that the bias takes the form of a minimum number of candidates that it is socially appropriate to vote for (one to four). Hence such a bias will only affect the reporting of the voters for whom their real turnout is pushed below the threshold. It is unlikely that the norm is strongly in favor of casting all four votes, since only approximately 55% of our respondents report voting for four candidates (with 34% voting for one to three, and 11% not voting). A minimum threshold of casting some of the potential votes might then shift the distribution of reported number candidates voted for between races towards more partial ballots (one to three candidates). However, the average number of candidates voted for is very similar between treatments (3.02 for control, 2.95 for Promise 1 and 3.00 for Promise 2), with no significant differences in the overall distributions between treatments (chi-squared test $p = 0.277$, all pairwise ranksum tests have $p > 0.20$). To the extent there is any difference, Promise 1 has more partial ballots than the Control treatment (37% versus 29%, proportions test $p = 0.04$), however Promise 2 also has a similar increase in partial ballots ($35\%, p = 0.09$). If social desirability bias with respect to turnout was distorting the distribution, it should only affect Promise 1, not Promise 2. Third, if the fake non-switches are mostly appearing in these partial ballots, we should see the largest treatment effect of Promise 1 in the city council race among partial ballots. However, if we split the sample between partial and complete ballots, the treatment effect is primarily coming from complete ballots ($\beta = -0.113$ for complete ballots, versus $\beta == 0.49$ for partial ballots).

Finally, we can look to additional data from our endline survey for evidence of treatment differences in turnout. Mullainathan and Washington (2009) show that, due to cognitive dissonance, the act of voting causes voters to more intensely identify with and have more polarized opinions of the candidates they voted for. Unfortunately, we don't have candidate level ratings in the endline survey, however we do ask subjects to recall who their favorite candidates were in the baseline survey.

[22] This assumption is consistent with models of promises and experimental evidence on lying. For example, Erat and Gneezy (2011) find that lying rates for "white lies" depend on the payoff consequences of the lie. Similarly, Miettinen (2013) models the guilt from breaking a promise as increasing in the payoff consequence of the promise violation.

[23] The overall level of social desirability bias could conceivably be higher in council races if vote selling was more common in less important races. However, as discussed previously, vote-selling is ubiquitous across all types of races in the Philippines, from Congress on down, and we have no qualitative or quantitative evidence to suggest that it is more common in less important races.

[24] If the Promise 1 treatment does have some true effect on reducing vote-switching in the higher-money races, then this estimate will be an upper bound of the true magnitude of social desirability bias. Also, if social desirability bias is not constant across electoral races but is larger in the more important races (mayor and vice-mayor), this estimate will also be an upper bound.

[25] An analogous assessment of the magnitude of social desirability bias is not possible for the Promise 2 treatment, because it is theoretically possible for that treatment to have a true positive effect on vote-switching, alongside any negative reporting bias due to social desirability effects.

If voting causes voters to attach more intensely to their candidate, then a turnout effect could cause a corresponding effect in recall. Specifically, if the true story is that switching rates are the same across treatments, but Promise 1 reduces voting, then we should expect to see lower recall in Promise 1. Additionally, consistent with the cognitive dissonance mechanism we do see that recall rates are significantly lower in all three races when participants report not voting (Difference of 9% for mayor, $p = 0.06$; difference of 35% for vice mayor, $p < 0.01$; difference of 62% for city council, $p < 0.01$). However, we don't see any difference in recall rates for any race by treatment - regressing recall on treatment dummies yields small, insignificant and directionally positive coefficients for the Promise 1 treatment (all $\beta$ are between 0.00 and 0.04, with $p > 0.30$). Therefore, while we cannot completely rule out the presence of social desirability bias for turnout in our data, the suggestive evidence we have is not consistent with such a bias being a major factor in our treatment effects.

### 6.5. Alternative mechanisms

The temptation model discussed above is presented as one potential mechanism that is consistent with our results, and as a way of being concrete about the potential impact of social desirability bias. Alternative mechanisms are certainly possible, but in order to explain our results an alternative mechanism would need to have two features. First, any alternate mechanism needs to explain why switching is more likely, and the effects of the promises are less positive (more negative), for races and candidates that offered more money to voters. Second, the alternative mechanism needs to explain why Promise 2 would be less effective than Promise 1, and increase the amount of vote-switching relative to the Control group. Therefore, an alternative mechanism needs to predict that (a) some voters will turn down money in the Control group, and (b) the mechanism by which money from candidates affects votes increases with the amount of money offered. The first feature explains how Promise 2 increases vote-switching, and the second feature explains the cross-race and cross-candidate differences.

Some potential mechanisms would have these features. For example, while we think that vote-buying primarily operates through reciprocity, to the extent that political brokers can exert coercive pressure on voters (c.f. Cruz, 2013), we would expect this to have similar effects. Voters would want to avoid such pressure—and hence may turn down money. Additionally, we would expect that brokers would exert more pressure on voters for more important races. Finally, one can imagine that voters may underestimate the amount of pressure they will face.

However, other mechanisms would not have the required features. For example, it is important to consider whether the treatments may affect voters' information on candidates, and thus their likelihood of vote-switching. Absent our treatments (in the control group), respondents would do some information-seeking between our baseline survey and the election, leading some to switch votes away from their previously-favored candidates. In the treatment groups, we might be concerned that they affect voters' information gathering in the following ways. Promise 1 ("don't take money") might cause respondents to reduce their information gathering, thus reducing their propensity to vote-switch. This could lead to the observed negative coefficient on the Promise 1 treatment. Promise 2 ("take money, vote conscience") could lead respondents to *increase* their information gathering, increasing their propensity to vote-switch. This would lead the coefficient on Promise 2 to be positive.

The first point is that there seems to be no obvious reason why Promise 1 should lead to less information gathering. If anything, one might think Promise 1 should lead to more information gathering than in the control group, since the promise is elicited in the context of asking voters to be more responsible about their voting. We do think it is more plausible to consider whether Promise 2 might have led to more information-seeking, because the promise to "vote one's conscience" may have led to such seeking. However, the pattern of results in Table 3 do not support this hypothesis. If the information-seeking hypothesis were empirically important, we would expect it to be more prominent where respondents had less information to begin with – the city council race, rather than the mayor and vice-mayor races. This would predict that the coefficient on Promise 2 would be more positive in the city council race than in the other races. This is the opposite of the pattern we find in the data. We conclude that the information-seeking hypothesis does not provide an explanation for the qualitative pattern of results across races in Table 3. By contrast, the behavioral model of vote-selling temptation we present in the paper does account for all the key patterns across coefficients.

Another alternate mechanism would be to assume that vote-selling operated through regular reciprocity (i.e. not temptation), but that voters' initial favorability reports incorporate both their true underlying preference and the anticipated monetary offers from each candidate. In this case Promise 2 would actually lead more voters to vote for their true preferred candidate, hence the apparent increase in switching would be an improvement in voting fidelity. However, in this model there would never be any apparent switching in the Control group (since voters have already factored in their vote-selling into their initial reports), and Promise 1 would also cause the same increase in apparent switching as Promise 2 (since eliminating the monetary payments also leads voters to vote for their true favorite).

## 7. Conclusion

We report the results of a randomized controlled trial of an anti-vote-selling intervention in the Philippines. We randomly assigned individual voters to treatments that invited them to make particular promises intended to reduce vote-selling. Across promises and across electoral races, we found unexpected patterns of impacts on a proxy measure of vote-selling. We outline a behavioral model of transactional electoral politics that makes sense of the results. In the model, selling one's vote is a temptation good, generating utility for the future self upon the vote-sale, but not for the present self who anticipates later selling his or her vote. We allow keeping or breaking promises to have utility consequences, so voters can use promises related to vote-selling as a commitment device. The model predicts that a promise not to take money from candidates can reduce vote-selling, but a different type of promise (to take vote-buying payments, but to nonetheless vote according to one's underlying candidate preferences) can have a smaller effect, and even possibly increase vote-selling, if voters are partially naïve about (underestimate) their vote-selling temptation. Our empirical results are consistent with the case wherein voters are partially naïve about their vote-selling temptation. The results rule out full sophistication as well as full naïveté about one's vote-selling temptation.

From a policy standpoint, our results reveal that exceedingly simple interventions — such as eliciting promises not to sell votes — can help reduce vote-selling. We estimate that a promise not to take money from candidates leads to a reduction in vote-switching (our proxy for vote-selling) of 10.9 percentage points (compared to a rate of 47.1 percent in the control group) in the electoral race that involved the smallest vote-buying payments (the city council race). Patterns in the results for other races indicate that this treatment effect estimate is likely to be minimally biased by social desirability effects. We find no evidence that promises help reduce vote-selling in the races (for mayor and vice-mayor) in which vote-buying payments are larger.

These results reveal that approaches from behavioral economics or psychology can help us understand important phenomena in political economy, such as vote-selling transactions. Future research would do well to incorporate the behavioral factors we have highlighted into theoretical and empirical analyses of transactional electoral politics, and of vote-selling in particular.

## Acknowledgements

We thank Joma Gonzalez (Innovations for Poverty Action) for unparalleled field management. Vibha Mehta for her contributions to the field work and analysis, and seminar participants at Georgia Institute of Technology, Simon Fraser University, University of British Columbia, the World Bank, Georgetown University, and the University of Sydney. This study was made possible by funding from MCubed program at the University of Michigan (Project ID No. 211).


## Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jdeveco.2017.10.012.

## References


Acemoglu, Daron, Robinson, James, Santos, Rafael, 2009. The Monopoly of Violence: Evidence from Colombia. NBER Working Paper No. 15578.
Alvarez, Michael, Nagler, Jonathan, 2000. A new approach for modelling strategic voting in multiparty elections. Br. J. Political Sci. 30, 57–75.
Ashraf, Nava, Karlan, Dean, Yin, Wesley, 2006. Tying odysseus to the mast: evidence from a commitment savings product in the Philippines. Q. J. Econ. 121 (2), 635–672.
Baland, Jean-Marie, Robinson, James, 2008. Land and power: theory and evidence from Chile. Am. Econ. Rev. 98 (5), 1737–1765.
Banerjee, Abhijit, Green, Donald, Pande, Rohini, 2012. Can Voters Be Primed to Choose Better Legislators? Evidence from Voter Campaigns in India. Working Paper.
Banerjee, Abhijit, Kumar, Selvan, Pande, Rohini, Su, Felix, 2011. Do Informed Voters Make Better Choices? Experimental Evidence from Urban India. Working Paper.
Banerjee, Abhijit, Mullainathan, Sendhil, 2010. The Shape of Temptation: Implications for the Economic Lives of the Poor. NBER Working Paper No. 15973.
Barabas, Jason, Jerit, Jennifer, 2010. Are survey experiments externally valid? Am. Political Sci. Rev. 104 (2), 226–242.
Beaman, Lori, Chattopadhyay, Raghabendra, Duflo, Esther, Pande, Rohini, Topalova, Petia, 2009. Powerful women: does exposure reduce bias? Q. J. Econ. 124 (4), 1497–1540.
Blaydes, Lisa, 2006. Who Votes in Authoritarian Elections and Why? Determinants of Voter Turnout in Contemporary Egypt. APSA 2006 Annual Meeting paper.
Bratton, Michael, 2008. Vote buying and violence in Nigerian election campaigns. Elect. Stud. 27 (4), 621–632.
Callahan, William, 2000. Pollwatching, Elections and Civil Society in Southeast Asia. Ashgate, Aldershot, Hampshire.
Charness, Gary, Dufwenberg, Martin, 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.
Chong, Alberto, Ana De La, O., Karlan, Dean, Wantchekon, Leonard, 2011. Looking beyond the Incumbent: the Effects of Exposing Corruption on Electoral Outcomes. NBER Working Paper No. 17679.
Cruz, Cesi, 2013. Social Networks and the Targeting of Illegal Electoral Strategies. Working Paper.
Cruz, Cesi, Keefer, Philip, Labonne, Julien, 2015. Incumbent Advantage, Voter Information, and Vote Buying. Working Paper.
Desposato, Scott, 2007. Elections for sale: the causes and consequences of vote buying. In: Boulder, Colorado: Lynne Rienner Chapter How Does Vote Buying Shape the Legislative Arena?, pp. 144–179.
Duflo, Esther, Kremer, Michael, Robinson, Jonathan, 2011. Nudging farmers to use fertilizer: theory and experimental evidence from Kenya. Am. Econ. Rev. 101 (6), 2350–2390.
Erat, Sanjiv, Gneezy, Uri, 2011. White lies. Manag. Sci. 58 (4), 723–733.
Ferraz, Claudio, Finan, Frederico, 2008. Exposing corrupt politicians: the effect of Brazil's publicly released audits on electoral outcomes. Q. J. Econ. 123 (2), 703–745.
Finan, Frederico, Schechter, Laura, 2012. Vote-buying and reciprocity. Econometrica 80 (2), 863–881.
Flores, Helen, Jaymalin, Mayen, Crisostomo, Sheila, 2013. Comelec: Vote Buying Rampant. The Philippine Star. http://www.philstar.com/headlines/2013/05/16/942666/comelec-vote-buying-rampant.
Fudenberg, Drew, Levine, David, 2006. A dual self model of impulse control. Am. Econ. Rev. 96 (5), 1449–1476.
Gaines, Brian, Kuklinski, James, Quirk, Paul, 2007. The logic of the survey experiment reexamined. Polit. Anal. 15 (1), 1–20.
Geronimo, Gian, 2013. Concerned Groups Launch Anti-vote-buying Campaign. GMA News online. February 13 http://www.gmanetwork.com/news/story/294800/news/nation/concerned-groupslaunch-anti-vote-buying-campaign.
Gine, Xavier, Mansuri, Ghazala, 2012. Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan. World Bank. Working Paper.
Giné Xavier, Goldberg Jessica, Silverman Dan and Yang Dean, Revising commitments: field evidence on adjustment of prior choices. Econ. J. Forthcoming.
Golden, Miriam, Tiwari, Devesh, 2009. Criminality and Malfeasance Among National Legislators in Contemporary India. APSA 2009 Annual Meeting paper.
Golden, Miriam, Kramon, Eric, Ofosu, George, 2014. Electoral Fraud and Biometric Identification Machine Failure in a Competitive Democracy. Working Paper.
Graziano, Luigi, 1973. Patron-client relationships in southern Italy. Eur. J. Political Res. 1 (1), 3–34.
Guiang, Alfred, 2013. Anti-vote-buying Mass Held at Q.C. Church to Promote Electoral Honesty and Integrity. Philippine Information Agency. March 18 http://ncr.pia.gov.ph/index.php?article=281363581025.
Gul, Faruk, Pesendorfer, Wolfgang, 2001. Temptation and self-control. Econometrica 69 (6), 1403–1435.
Hicken, Allen, 2009. Building Party Systems in Developing Democracies. Cambridge University Press, New York, New York.
Hicken, Allen, 2011. Clientelism. Annu. Rev. Political Sci. 14, 289–310.
Hicken Allen, Aspinall Edward, Weiss Meredith (Eds.), Money Politics in the Philippines Forthcoming.
Hicken, Allen, Simmons, Joel, 2008. The personal vote and the efficacy of education spending. Am. J. Political Sci. 52 (1), 109–124.
Hicken, Allen, Leider, Stephen, Ravanilla, Nico, Yang, Dean, 2015. Measuring vote-selling: field evidence from the Philippines. Am. Econ. Rev. Pap. Proc. 105 (5), 352–356.
Kaur, Supreet, Kremer, Michael, Mullainathan, Sendhil, 2017. Self-control at work. J. Political Econ. 123 (6), 1227–1277.
Keefer, Philip, 2006. Programmatic parties: where do they come from and do they matter? In: Presented at the Annual Meeting of the American Political Science Association, Aug. 31Sep. 3, Philadelpia.
Keefer, Philip, 2007. Clientelism, credibility, and the policy choices of young democracies. Am. J. Political Sci. 51 (4), 804–821.
Kessler, Judd, Leider, Stephen, 2012. Norms and contracting. Manag. Sci. 58 (1), 62–77.
Khemani, Stuti, 2013. Buying Votes versus Supplying Public Services: Political Incentives to Under-invest in Pro-poor Policies. World Bank. Working Paper.
Kitschelt, Herbert, Hawkins, Kirk, Pablo Luna, Juan, Rosas, Guillermo, Zechmeister, Elizabeth, 2010. Latin American Party Systems. Cambridge University Press, Cambridge, United Kingdom.
Kitschelt, Herbert, Wilkinson, Steven, 2007. Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition. Cambridge University Press, New York. chapter Citizen-Politician Linkages: An Introduction.
Krupka, Erin, Leider, Stephen, Jiang, Ming, 2013. A Meeting of the Minds: Informal Agreements and Social Norms. University of Michigan. Working Paper.
Laibson, David, 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112 (2), 443–477.
Lyne, Mona, 2007. Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition. Cambridge University Press, Cambridge, United Kingdom, pp. 206–226. chapter Rethinking Economics and Institutions: the voters dilemma and democratic accountability.
Miettinen, Topi, 2013. Promises and conventions: an approach to preplay agreements. Games Econ. Behav. 80, 68–84.
Mullainathan, Sendhil, Washington, Ebonya, 2009. Sticking with your vote: cognitive dissonance and political attitudes. Am. Econ. J. Appl. Econ. 1, 86–111.
Olken, Ben, Pande, Rohini, 2011. Can informed voters enforce better Governance? Experiments in low income democracies. Annu. Rev. Econ. 3 (1), 215–237.
Quijano, Therene, 2013. Discreet Vote-buying Cases Reported in Dumaguete. Rappler. http://www.rappler.com/nation/politics/elections-2013/local-races/the-visayas/28637-discreet-vote-buying-cases-reported-in-dumaguete.
Schaffer, Frederic, 2005. Clean Elections and the Great Unwashed: Educating Voters in the Philippines. Occasional Paper 21. School of Social Science, IAS.
Scott, James, 1969. Corruption, machine politics, and political change. Am. Political Sci. Rev. 63 (4), 1142–1158.
Shu, Lisa, Mazar, Nina, Gino, Francesca, Ariely, Dan, Bazerman, Max, 2012. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proc. Natl. Acad. Sci. 109 (38), 15197–15200.
Stokes, Susan, 2005. Perverse accountability: a formal model of machine politics with evidence from Argentina. Am. Political Sci. Rev. 99 (3), 315–325.
Stokes, Susan, Dunning, Thad, Nazareno, Marcelo, Brusco, Valeria, 2013. Brokers, Voters, and Clientelism: the Puzzle of Distributive Politics. Cambridge University Press, Cambridge, United Kingdom.
Vanberg, Christoph, 2008. Why do people keep their promises? An experimental test of two explanations. Econometrica 76 (6), 1467–1480.
Vicente, Pedro, 2014. Is vote buying Effective? Evidence from a field experiment in west Africa. Econ. J. 124 (574), 356–387.
Wantchekon, Leonard, 2003. Clientelism and voting behavior: evidence from a field experiment in Benin. World Polit. 55 (3), 399–422.