



Could vote buying be socially desirable? Exploratory analyses of a ‘failed’ list experiment

Sophia Hatz¹ · Hanne Fjelde¹ · David Randahl¹

Accepted: 25 August 2023
© The Author(s) 2023

Abstract

List experiments encourage survey respondents to report sensitive opinions they may prefer not to reveal. But, studies sometimes find that respondents admit more readily to sensitive opinions when asked directly. Often this over-reporting is viewed as a design failure, attributable to inattentiveness or other nonstrategic error. This paper conducts an exploratory analysis of such a ‘failed’ list experiment measuring vote buying in the 2019 Nigerian presidential election. We take this opportunity to explore our assumptions about vote buying. Although vote buying is illegal and stigmatized in many countries, a significant literature links such exchanges to patron-client networks that are imbued with trust, reciprocity and long-standing benefits, which might create incentives for individuals to claim having been offered to participate in vote buying. Submitting our data to a series of tests of design, we find that over-reporting is strategic: respondents intentionally reveal vote buying and it’s likely that those who reveal vote buying have in fact being offered to participate in vote buying. Considering reasons for over-reporting such as social desirability and network benefits, and the strategic nature of over-reporting, we suggest that “design failure” is not the only possible conclusion from unexpected list experiment results. With this paper we show that our theoretical assumptions about sensitivity bias affect the conclusions we can draw from a list experiment.

Keywords Measurement · List experiment · Social desirability bias · Vote buying

1 Introduction

Decades of survey research has established that individuals often do not provide truthful answers when asked sensitive questions (Blair et al. 2020; Glynn 2013; Tourangeau and Yan 2007; Fisher 1993; Rinken et al. 2021). This occurs across diverse types of sensitive

✉ Sophia Hatz
sophia.hatz@pcr.uu.se

Hanne Fjelde
hanne.fjelde@pcr.uu.se

David Randahl
david.randahl@pcr.uu.se

¹ Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden

topics, from drug-use to voting behavior, and is a problem recognized in nearly all social science literatures (Fisher 1993).

The problem, known as sensitivity bias,¹ has been a significant challenge in the study of vote buying. Vote buying is a particular form of clientelism which involves the exchange of money or material incentives in return for votes (Gonzalez-Ocantos et al. 2012; Stokes 2007). In many countries, vote buying is both illegal and socially stigmatised (Stokes 2007), making it a sensitive topic to ask individuals about. By now, scholars take the under-reporting of vote buying as given, and few scholars would ask about vote buying without using a technique for the measurement of sensitive topics. In particular, many scholars have turned to list experiments— a measurement technique which offers surveys respondents confidentiality in order to encourage truthful reporting.

Following common practices, we employed a list experiment within a nationally representative survey in order to measure vote buying in the 2019 Nigerian general elections.² We also measured vote buying using a direct survey question. When comparing the list experiment estimate of vote buying and the direct measure, we find that the reported prevalence of vote buying is 22 percentage points higher in the direct survey question. Often this result—the *over-reporting* of a sensitive behavior³— is viewed as a list experiment “failure” (e.g., Kramon and Weghorst 2019). This kind of result is not uncommon, occurring in contexts as diverse as shoplifting in Japan (Tsuchiya et al. 2007) and support for same-sex marriage in the U.S. (Lax et al. 2016). Yet, the over-reporting of a sensitive behavior is not well understood, and has contributed to a file-drawer problem in the context of list experiments (Blair et al. 2020; Kramon and Weghorst 2019; Gelman 2014; Castro Cornejo and Beltrán 2020, 2022).

In this paper, we conduct an exploratory analysis of our ‘failed’ list experiment. Specifically, we take the opportunity to explore our assumptions about sensitivity bias in the context of vote buying. Prior research identifies several reasons why individuals might prefer to conceal vote buying, such as fear of social stigma and legal sanctions. Yet, the literature on clientelism also describes vote buying as a mutually beneficial patron-client exchange, imbued with trust, reciprocity and potential for long-term, tangible rewards. This suggests there are incentives—such as social desirability and anticipation of network benefits—for individuals to claim having been subject to vote buying offers. If this is the case, then *not* having participated in vote buying may be considered the sensitive response in a survey question, and the expected outcome of sensitivity bias is over-reporting.

In our exploratory analysis, we show how assumptions about polarity (what the sensitive response is and whether under- or over-reporting is expected) affect the conclusions we draw from a list experiment.⁴ In particular, we show how the interpretation of tests meant

¹ This is also referred to as response bias or social desirability bias (Krumpal 2013; Fisher 1993). We use the broader term introduced in Blair et al. (2020) which encompasses misreporting for a variety of reasons not limited to social desirability.

² Vote buying can be defined as the “offering [of] particularistic material rewards to individuals or families at election time” (Schaffer 2007). When studying the prevalence of vote buying, we focus on the targeting of voters with offers of money, while remaining agnostic about whether the offer is accepted or not and whether the voter actually complies at the polling booth (e.g., Mares and Young 2019). When we use the term “participated in vote buying” in this study we refer to being targeted by vote buying offers.

³ Throughout this paper, we refer to a higher reported prevalence in a direct measure compared to a list experiment as “over-reporting”.

⁴ In supplementary material, Blair et al. (2020) use the term “polarity” to refer to whether under- or over-reporting is expected.

to diagnose design failure also depend on polarity assumptions. First, following standard methodology, we find no evidence of failure due to violations of the “no design effects” and “no liars” assumptions of list experiments (Blair and Imai 2012). Respondents do not appear to adjust their responses to the list experiment depending on the inclusion of the vote buying item (no design effects). Formal tests for “floor”- and “ceiling”-liars suggest that respondents do not conceal vote buying in the list experiment. However, we find evidence that respondents intentionally reveal vote buying in the list experiment, which could be interpreted as evidence of sensitivity bias under the assumption that answering “no” is the sensitive response.

Second, we examine the extent of uniform error and top-biased error in the list experiment. These are examples of “nonstrategic” measurement error; misreporting due to respondent inattentiveness or satisficing (Ahlquist 2018; Blair et al. 2019). Here, we find that respondents may choose random responses in the list experiment as a way of concealing whether or not they have participated in vote buying. This suggests that evidence of uniform error can be interpreted as strategic misreporting. We also find that top-coding – which implies revealing vote buying in the list experiment– is very prevalent, and note that this can be viewed as strategic under the assumption that answering “no” is sensitive. Further, when modeling the correlation between non-sensitive items and the latent sensitive trait, we find that a large proportion of top-coders are likely to truly have participated in vote buying. Taken together, the results of our tests of design suggest that the over-reporting we observe is strategic in nature. This is consistent with theories of vote buying which emphasise its positive connotations, such as social desirability and network benefits. We conclude that ‘design failure’ is not the only reasonable interpretation of unexpected list experiment results.

While there is a substantial amount of prior research employing list experiments to measure vote buying, this paper differs in several specific ways. First, we connect the literature on measurement and the literature on clientelism in order to theorise about the likely reasons for over-reporting vote buying. While the measurement literature identifies diverse sources of sensitivity bias, which manifest in different polarities (Blair et al. 2020; Krumpal 2013), prior research on vote buying has focused almost exclusively on motivations for under-reporting. Second, we apply recent methodological advancements in list experiment diagnostic tests to our list experiment under both assumptions of polarity. Existing publications, in contrast, only present findings under the assumption that vote buying will be under-reported.

The findings have implications for our understandings of clientelism and electoral democracy. While vote-buying has perverse effects on democracy by distorting accountability and violating the fairness of democratic procedures (Stokes 2005), our findings suggest that, in Nigeria, vote buying might in fact be considered legitimate and socially desirable by a significant segment of the population. If over-reporting testifies to norms that not only tolerate, but even encourage vote buying at the level of the individual voter, it might explain the pervasiveness of vote buying practices long after democratic transitions.

The paper also has broader implications. For many sensitive topics, scholars have strong uni-directional assumptions of polarity (Gelman 2014). For example, we expect under-reporting for drug-use and over-reporting for voter turnout. We show how these kinds of assumptions guide what is interpreted as sensitivity bias and what is interpreted as design failure. The implication is that, in some situations, incorrect assumptions could lead to the interpretation of unexpected results as design failures, and to the dismissal of potentially valuable empirical findings. This is relevant for any field of research which relies on self-reporting for the measurement sensitive attitudes and behaviors.

2 Sensitivity bias, list experiments and vote buying

Sensitivity bias is a well-known problem in survey research. The problem is that when survey respondents misreport, survey questions provide inaccurate estimates of aggregate attitudes and behaviors. Because the error in reporting is not random (certain individuals tend to misreport), this creates systematic bias in analyses of the causes and consequences of sensitive topics.

Sensitivity bias is a significant challenge in the social sciences, which often aims to study controversial, taboo, private or otherwise sensitive topics. In the case of vote buying, existing studies provides several reasons why survey respondents may want to avoid reporting participation in vote buying. To begin with, vote buying is illegal in many countries, so respondents may fear legal sanctions (Gonzalez-Ocantos et al. 2012; Mares et al. 2018; Blair et al. 2020; Mares and Young 2019; Stokes 2007). Vote buying may also be socially stigmatised because it violates norms of democratic conduct (Kramon 2016; Corstange 2018), and because of its association with low socio-economic standing (Cammett 2014; Corstange 2018). The existence of social norms which make vote buying socially undesirable also make self-preservation a potential motivation for concealing vote buying: respondents may seek to preserve a positive self-image by denying to themselves that they have violated what they consider to be prevailing norms.

In survey research, these kinds of arguments have motivated the use of list experiments, a technique designed to provide accurate measures of sensitive topics (Blair et al. 2020). List experiments encourage truthful reports of sensitive behaviors by offering respondents greater privacy. Studies using list experiments often also include direct measures of the sensitive topic: standard survey questions without additional privacy. By comparing list experiment estimates to estimates from direct survey questions, scholars are able to detect whether (in the aggregate) respondents answer differently when afforded additional privacy, compared to when asked directly. The difference in estimated prevalence rates is then generally interpreted as evidence of sensitivity bias (Blair and Imai 2012).

In the study of vote buying, sensitivity bias appears common, large in magnitude and consistent in direction. In a meta-analysis of 19 survey studies Blair et al. (2020) conclude that vote buying tends to be under-reported by an average of 8 percentage points. Based both on theory and accumulated evidence, many scholars now assume that vote buying will be under-reported in direct measures (e.g., Blair et al. (2020), Carkoglu and Aytaç (2015, p. 548), Kiewiet De Jonge (2015, p. 720)). However, this prevailing view is partially maintained because studies which do not find evidence of under-reporting tend to be dismissed. It is actually not uncommon that list experiments on vote buying fail to reveal under-reporting (c.f. Castro Cornejo and Beltrán (2020, p. 224)). Returning to Blair et al. (2020)'s meta analysis: 7 out of 19 studies found significant under-reporting, but the remaining 12 studies found no significant discrepancies across a direct measure and a list experiment. Further, 4 of these 12 studies revealed marginally higher prevalence levels in a direct measure, although this over-reporting is not statistically significant. However, as list experiment practitioners have noted, there is a tendency for scholars with null or unexpected list experiment results not to pursue publication, ultimately leading to a file-drawer problem which may slant evidence from meta-analyses (Blair et al. (2020, p. 1306), Kramon and Weghorst (2019, p. 239), Castro Cornejo and Beltrán (2020, p. 224), (Castro Cornejo and Beltrán 2022, p. 10), Gelman (2014)). In the context of vote buying in particular, the assumption that vote buying should be under-reported may cause a kind of confirmation bias where evidence suggesting no sensitivity bias or over-reporting is dismissed and

evidence suggesting under-reporting is over-emphasized.⁵ Even those studies which are transparent about discovering significant evidence of over-reporting of vote buying tend to attribute this to respondent error (Castro Cornejo and Beltrán 2020, 2022), design failure or “list experiment breakdown” (Kramon and Weghorst 2019; Kerr 2018).

We consider the tendency to interpret over-reporting as evidence of design failure to be problematic, since what is deemed as a failure depends on our assumptions about the *polarity* of the sensitive topic: what the sensitive response to the question is, and accordingly, whether under- or over-reporting is the expected indication of sensitivity bias. In the case of vote buying, scholars assume that answering affirmatively when asked about vote buying is the sensitive response; so vote buying will be under-reported in direct survey questions. Uni-directional assumptions are not limited to vote buying. For example, Blair et al. (2020) review four political science literatures, identifying for each topic whether under- or over-reporting is expected (see also Gelman (2014)). These uni-directional assumptions seem overly restrictive, considering that the measurement literature identifies a number of different sources of sensitivity bias, which could lead to different polarities for a given topic (Blair et al. 2020; Krumpal 2013).

Based on research on clientelism, we see reasons to reconsider our strong assumptions of polarity when measuring vote buying. Much ethnographic and comparative research suggests that vote buying might not always carry negative social stigma, but also convey benefits, even beyond what is entailed in the specific transaction. To begin with, a vote buying exchange might be embedded in social norms—such as reciprocity and solidarity—which make vote buying a socially desirable behavior. Whereas vote buying within survey research is measured as an isolated exchange close to or on election day, the networks invoked in this form of electoral mobilization oftentimes reflect long-standing relationships between patrons and local communities through which selective benefits such as gifts, services, favors and protection are exchanged for political loyalty (Lawson and Greene 2014). Such patron-client relationships often build on norms of reciprocity in a “moral economy of exchange” (Scott 1972). Research suggests that actors involved in vote-buying transactions (also referred to as election-related gift-giving or favour-rendering) often see them as mutually beneficial, and part of an everyday “problem-solving-network” (e.g., Gonzalez Ocantos et al. 2014; Kiewiet De Jonge 2015; Schaffer 2007). From the perspective of the political candidates, wealth and willingness to give gifts can be associated with authority and legitimacy, and the disbursement of electoral handouts as a sign of civic virtue (Chabal and Daloz 1999; Nugent 2007; Gadjanova 2017). These features also serve to normalize the distribution of electoral handouts amongst voters, and reduces the social stigma associated with it (see e.g., Gadjanova 2017; Nugent 2007; van de Walle 2007). Feelings of obligation, gratitude and norms of reciprocity have been used to explain voters’ compliance

⁵ To give an example: the results from a set of 10 list experiments conducted in Latin America show significant under-reporting of vote buying in 3 studies and no significant difference in 7 studies (Figure 1, p. 720, Kiewiet De Jonge 2015). Of the 7 studies with insignificant differences, 3 of them suggest lower prevalences in the direct measure (under-reporting), 3 of them reveal higher prevalence levels in direct measure (over-reporting) and 1 yields nearly equal prevalence levels (no sensitivity bias). But, Kiewiet De Jonge (2015) measures sensitivity bias as an absolute value, adjusting estimates indicating over-reporting to 0, since higher prevalence rates in the direct measure “are not theoretically possible” (2015, p. 720). With this adjustment, Kiewiet De Jonge (2015) reports evidence of “widespread” under-reporting, suggesting that 6 studies revealed social desirability bias and that under-reporting varies from 22 percentage points to essentially 0 (even though differences were not significant in 3 of these 6 studies, and 3 of the 0 estimates were indicative of over-reporting) (2015, p. 710–11, 720).

with vote-buying offers - even in the presence of ballot secrecy and absent strong monitoring mechanisms (Lawson and Greene 2014; Finan and Schechter 2012). This discussion confirms social desirability and self-presentation as relevant sources of sensitivity bias, but in this case implying that *not* having participated in vote-buying is the sensitive response, and that vote buying will be over-reported when respondents are asked directly.

To summarise our discussion, Table 1 lists the possible sources of sensitivity bias we have discussed. For each motivation, we list the corresponding polarity assumptions: what the sensitive response to a survey question is and the expected form of sensitivity bias. Blair et al. (2020, p.1300) provides a similar table for vote buying and three other sensitive topics. For vote buying, the motivations Blair et al. (2020) list all result in respondents preferring to answer “no”, with the implication that vote buying will be under-reported. We have additionally theorised about possible motivations which could lead respondents to prefer to answer affirmatively, such as those Blair et al. (2020) and Krumpal (2013) identify for other sensitive topics.

3 Survey data and design

For our empirical analysis, we draw on an original, nationally representative survey of 2400 Nigerian citizens fielded just after the 2019 general elections.⁶ Nigeria represents a suitable case to explore assumptions about the sensitivity of vote buying. On one hand, vote buying is illegal in Nigeria and recent laws trying to curtail its prevalence has highlighted its unlawful and undemocratic aspects (Obe 2019). On the other hand, there is also research suggesting that vote buying and vote selling has become widely accepted norm in Nigeria (Sakariyau et al. 2015) and is deeply intertwined with the nature of electoral politics in the country (Olaniyan 2020). Obe (2019, p. 13), for example remarks on the peculiar honesty of those engaged in vote-selling in Nigeria, noting that many voters in the 2019 elections expected money to show up and openly asked candidates what they were willing to pay. Systematic data from earlier electoral rounds in Nigeria testify to ambiguous popular attitudes to vote buying. Bratton (2008) reports that amongst respondents surveyed by the Afrobarometer after the 2007 election, barely half (49%) thought it was “wrong and punishable” for a Nigerian voter to “accept money in return for a vote”, whereas the other half was willing to excuse participation in a vote buying transaction as “wrong but understandable” (35%) or “not wrong at all” (10%). Comparative studies show that about 28% of Nigerians admit to vote buying in direct survey questions, well above the median rate of 15% (Carkoglu and Aytac 2015, p. 550), which is another indication that vote buying may be less sensitive or have positive connotations in Nigeria relative to other countries. Our survey asked about political attitudes and election day experiences, and included both a list experiment measure and a direct measure of vote buying.⁷

In the list experiment, respondents were read a list of statements about things they could have experienced during the 2019 election campaign, and were asked to state *how many* of those things they had experienced, but not *which* ones. The control group was read a list of four non-sensitive control items (List A in Table 2), while the treatment group was

⁶ More detail about the survey design, sampling and implementation is found in section A of the online appendix.

⁷ The survey and list experiment were approved by the Swedish Ethical Review Board on 2018-11-07 (ID no 2018/409). A discussion of ethical considerations is included in section A.3 of the online appendix.

Table 1 Possible sources of sensitivity bias in vote buying

Motivation for misreporting	Polarity assumptions	
	Sensitive response	Expected sensitivity bias
Fear of legal sanctions	Yes	Under-reporting
Social desirability (fear of stigma)	Yes	Under-reporting
Self-presentation (violation of norms)	Yes	Under-reporting
Network benefits	No	Over-reporting
Social desirability (desire to conform)	No	Over-reporting
Self-presentation (conformity with norms)	No	Over-reporting

Note: This table lists possible motivations for misreporting vote buying in a survey. For each motivation, we list the corresponding polarity assumptions: the expected sensitive response to a survey question and the expected form of sensitivity bias

read the same list with the addition of a sensitive item on vote buying (List B in Table 2). By virtue of random assignment, the list experiment measure of vote buying is the difference in the mean count of items across treatment and control groups. The list experiment difference-in-means is an aggregate measure; whether respondents reported vote buying in the list experiment is not observable at the individual level in order to protect respondents' privacy.

To avoid potential priming effects, we placed the direct measure of vote buying later in the survey after a series of unrelated questions.⁸ In the direct measure, respondents were read a list of statements about ways political actors may have tried to persuade them to vote in a particular way during the 2019 elections, and were asked to state which ones they experienced. The direct measure of vote buying is the proportion of respondents who answered affirmatively to the vote buying item.

When designing the list experiment and direct measure, we strove to ensure that the two questions measure the same quantity, and that this quantity corresponds well to conceptualisations of vote buying. First, both questions ask about an offer of material incentives “to vote in a particular way”. Conceptually, it is important to include a phrase relating to expectations of an exchange in order to capture the *quid pro quo* nature of vote buying offers (Stokes 2005) and in order to ensure that respondents are reporting a clientelistic exchange (rather than a non-clientelistic exchange, such as the receipt of campaign gifts) (Castro Cornejo and Beltrán 2020, 2022; Kiewiet De Jonge 2015). Second, both questions ask about the receipt of an *offer* rather than whether respondents *accepted* the offer. This helps us to capture the socially desirable aspects of vote buying as a clientelistic exchange; the receipt of an offer is likely to be socially desirable in a clientelistic context, while not receiving an offer could be an undesirable signal of exclusion from patron-client networks. Third, both questions ask about respondents' personal experiences. This is important, as prior research has shown that individuals report vote buying differently when asked about themselves or others (Gonzalez-Ocantos et al. 2012). Fourth, both questions specify an offer of money as prior research has shown that respondents answer differently when asked about money

⁸ While we are unable to test for priming effects directly, we show in online appendix B.2 that the distribution of responses to the direct question are effectively the same across respondents assigned to treatment and control in the list experiment.

Table 2 List experiment statements

List A (control)	List B (treatment)
1. Politicians put up posters or signs in the area where you live	1. Politicians put up posters or signs in the area where you live
2. You read the newspaper almost every day to learn about the campaign	2. You read the newspaper almost every day to learn about the campaign
3. You met a politician personally to discuss his or her candidacy	3. You were offered money from a party or politician to vote in a particular way
4. You discussed the campaign with friends or family	4. You met a politician personally to discuss his or her candidacy
	5. You discussed the campaign with friends or family

as opposed to other kinds of non-monetary compensation (Kiewiet De Jonge 2015; Castro Cornejo and Beltrán 2020). Our list experiment and direct measures of vote buying are similar to those in Kramon (2016), Gonzalez-Ocantos et al. (2012) and Kiewiet De Jonge (2015).

In addition to our measures of vote buying, we used survey questions on demographic characteristics to construct control variables for our analysis. Drawing on prior related research, we control for age, gender, college education and employment (e.g., Gonzalez-Ocantos et al. 2012; Carkoglu and Aytac 2015; Castro Cornejo and Beltrán 2020). We additionally include an indicator of whether the respondent is a registered voter, since registered- and non-registered voters received different survey questions.⁹

4 Results

Following the standard approach to detect sensitivity bias (e.g., Blair et al. 2020; Aro-now et al. 2015; Blair and Imai 2012), we compare the proportion of respondents who reported vote buying across the list experiment and the direct measure in our survey. The first plot in Fig. 1 shows reported vote buying in each of the two measures. We plot mean levels of the direct measure over the distribution of responses, to remind us that individual responses are observable in the direct measure, while in the list experiment vote buying can only be calculated in the aggregate. The bottom plot in Fig. 1 shows the estimated difference in reported vote buying comparing the direct measure to the list experiment.

As Fig. 1 shows, we find that survey respondents report higher levels of vote buying in the direct measure compared to the list experiment. This difference is both substantively large, 22 percentage points, and significant at the 95% confidence level.¹⁰

⁹ Section C of the online appendix includes the full wording of the survey questions.

¹⁰ Summary statistics for the list experiment, the direct measure of vote buying, and some demographic characteristics, as well as the raw scores for the list experiment in the treatment and control groups are available in online appendix B.1.

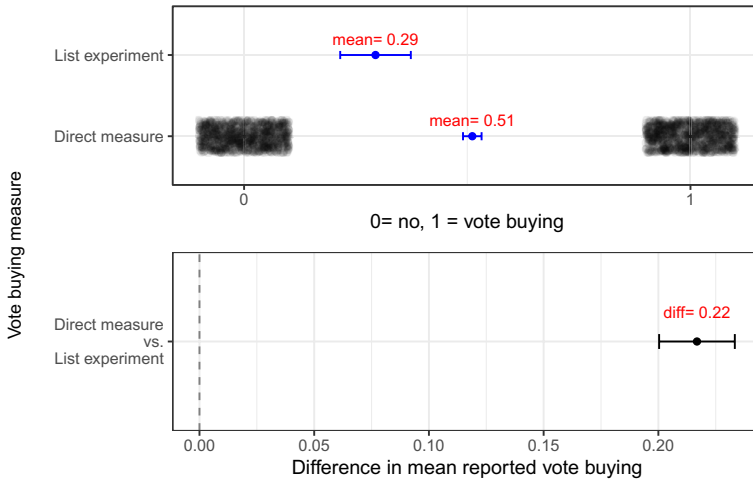


Fig. 1 Reported vote buying in the list experiment and direct measure. The top plot shows the proportion of respondents reporting vote buying in each measure. For the list experiment, the proportion of respondents who reported vote buying is the difference-in-means estimator (DiM), calculated via linear least squares. For the direct measure, the proportion of respondents who reported vote buying is the mean, calculated via robust linear regression. The bottom plot shows the estimated difference in reported vote buying. A 95% confidence interval for this difference is calculated via a two-tailed difference of proportions test

4.1 Test of design

Given that we have observed a significant difference across alternative measures of vote buying, our exploratory analysis aims to assess whether this discrepancy is truly evidence of sensitivity bias in a situation where vote buying is socially desirable, or whether the over-reporting we observe indicates a failed list experiment. Essentially, this requires distinguishing between strategic misreporting (as implied by conceptualisations of sensitivity bias) and nonstrategic measurement error. Nonstrategic measurement error can be the result of respondent inattentiveness/misunderstanding, or due to flaws in the survey design such as differences in question wording or priming effects (Ahlquist 2018; Blair et al. 2019; Kuhn and Vivyan 2021; Rimbau and Ostwald 2021).

Determining whether or not a list experiment has ‘failed’ is traditionally done by using a number of post-hoc tests to assess if the data conform to a number of crucial list experiment assumptions. However, as we show below, the interpretation of some of these tests is dependent on which answer we assume to be the sensitive response, i.e. the assumption of polarity. In the sections that follow we apply a number of different tests in attempt to assess whether vote buying is actually socially desirable, as the simple comparison of means in Fig. 1 suggests, or whether the apparent over-reporting of vote buying is attributable to violations of list experiment assumptions in the form of nonstrategic measurement error.

4.1.1 No design effects and no liars

Two essential assumptions must be met in order to trust that the proportion of respondents who admit to the sensitive behavior in a list experiment reflects the true prevalence of the sensitive behavior in the larger population. The “No design effects” assumption

states that the inclusion of the sensitive item in the list does not affect the sum of responses to the control items. The “No Liars” assumption in the design of list experiments states that respondents will not lie about the sensitive item in the list experiment (Blair and Imai 2012, p.51-2). These assumptions are closely related, and are often discussed and tested jointly (Blair and Imai 2012; Aronow et al. 2015).

Design effects can occur if the inclusion of the sensitive item in the treatment group causes respondents to adjust the total number of control items they report up (positive design effects) or down (negative design effects). As an initial diagnostic, we examine whether respondents in the treatment and group control groups answer the control items differently.

For this, we start with the methodology proposed by Blair and Imai (2012) where we use the distribution of answers to the list experiment in the treatment and control groups to estimate the proportion of respondent types in the control group defined by two latent variables: the total number of control items reported Y_i (where Y ranges from 0 to $J=4$), and their answer to the sensitive item Z_i (coded 1 for yes, 0 for no). This is the distribution of responses in the control group which would be needed to replicate the distribution of answers in the treatment group. If any of the estimated proportions turn out to be negative, this indicates that less than 0% or more than 100% of respondents in that group would need to answer “yes” to the vote buying item in order to mimic the distribution of treatment group. If such negative proportions exist we can apply a formal test to assess the likelihood of observing such negative proportions.¹¹

In Table 3 we calculate the estimated proportion of each respondent type in the control group, defined by their latent count of control items ($Y(0) \in 0, \dots, 4$) and their latent answer to the sensitive item ($Z \in 0, 1$), as well as the estimated proportions of each respondent type who truly have participated in vote buying ($Z^* = 1$), which are needed to replicate the distribution of answers in the treatment group. As no negative proportions are observed in the first two columns we have no obvious violations of the no design effects assumption. However, the third column shows that the proportion of respondents in the control group who would need to answer affirmatively to the sensitive item in order to mimic the distribution of responses in the treatment group is substantively higher among respondents who reported four control items (74%) compared to respondents who reported three or fewer control items (0-30%). This pattern could be an indication that there are indeed design effects at play although these are not large enough to be detected by Blair and Imai’s (2012) test. However, this pattern could also reflect a correlation between the control items and the sensitive item (not a violation of the list experiment assumption) or the presence of top-biased liars, discussed below.

To further probe violations of the “No liars” assumption, we look for “floor” and “ceiling” effects in the survey data. Floor and ceiling effects are a well-known limitation of list experiments, as these are situations in which respondents do not answer truthfully to the sensitive item (Blair and Imai 2012; Ahlquist 2018). Although the list experiment offers confidentiality and should encourage truthful responses, in fact there are three situations in which the list experiment cannot offer confidentiality: i) if respondents in the treatment group have experienced all items in the list (Blair and Imai 2012; Ahlquist 2018), ii) if

¹¹ If such negative proportions do not exist, the test will fail to reject the null hypothesis with a p-value of 1. This implies that only large design effects can be detected by this test. A failure to reject the null should therefore not be taken as evidence to support the assumption of no design effects (Blair and Imai 2012, p. 65).

Table 3 Estimated proportion of respondent types, by latent list experiment and sensitive item responses

List experiment control items	$Z_i = 0$	$Z_i = 1$	$Z_i^* = 1$
$Y_i(0)=0$	0.01	0.00	0.00
$Y_i(0)=1$	0.24	0.03	0.11
$Y_i(0)=2$	0.24	0.08	0.25
$Y_i(0)=3$	0.18	0.08	0.30
$Y_i(0)=4$	0.04	0.11	0.74

This table shows the estimated proportion of each respondent type in the control group, defined by their latent count of control items ($Y(0) \in 0, \dots, 4$) and their latent answer to the sensitive item ($Z \in 0, 1$). We also show the estimated proportions of each respondent type with the latent sensitive trait ($Z^* = 1$) which are needed to replicate the distribution of answers in the treatment group.

respondents in the treatment group have experienced zero items in the list (Ahlquist 2018) or iii) if the control items are so controversial or uncommon that virtually everyone is expected to answer 0 to the control items (Blair and Imai 2012). In situation i) respondents may choose to answer 4 instead of 5 in order to not confess to the sensitive behavior – this is a “ceiling effect”. In situation ii) respondents may answer 1 instead of 0, and in situation iii) respondents may prefer to answer 0 instead of 1– these are both referred to as “floor effects”.

Floor and ceiling effects are considered examples of “strategic” measurement error, since respondents are intentionally avoiding revealing the truth (Ahlquist 2018; Rimbau and Ostwald 2021). So, in the specific case of the “no liars” assumption, a violation of the assumption does not correspond to design failure, rather it is a clue that misreporting in the list experiment might be attributable to sensitivity bias.

Floor- and ceiling effects would in general be expected to show up as negative proportions of respondent types in Table 3. As no such negative proportions are seen we have no obvious floor- or ceiling effects. Instead, the fact that the estimated proportion of respondents who have participated in vote buying is much higher among the individuals who reported four control items (74%) compared to the individuals who reported three control items (30%) suggests that the respondents do not try to avoid revealing vote buying by avoiding reporting the maximum number of items.

Yet, this interpretation depends on our assumption of polarity. If we assume that answering affirmatively to vote buying is stigmatised, then it would indeed be strategic for respondents to lie and answer 4 instead of the truthful 5 items in the list experiment to avoid confessing to the sensitive behavior. However, if we instead assume vote buying is socially desirable, it could instead be seen as strategic to answer 5 items instead of 4. If this is the case then the large estimated proportion of respondent types who participated in vote buying and answer affirmatively to the maximum number of control items might instead be seen as an indication of the presence of ceiling-liars, i.e. respondents who intentionally select the maximum value on the list experiment in order to reveal that they engage in a

socially desirable behavior.¹² The point is: our assumptions of polarity determine which respondent types are considered floor- and ceiling liars.¹³

A further implication is that formal tests for floor and ceiling effects could fail to detect sensitivity bias if our assumptions about polarity are wrong. To illustrate this further, we use two maximum likelihood (ML) estimators proposed by Blair and Imai (2012) which model floor and ceiling effects respectively. These models fit a separate sub-model to the data which models the likelihood of ceiling- and floor effects among respondents. The results support the no liars assumption: the estimated proportion of respondents who are floor-liars and ceiling-liars are each practically zero.¹⁴

However, these models are looking for floor- and ceiling-liars as defined by Blair and Imai (2012), which do not correspond to strategic misreporting under the assumption that answering negatively is the sensitive response. If vote buying is socially desirable, evidence of floor and ceiling effects might instead be interpreted as evidence of nonstrategic error, since these behaviors seem irrational. Following this logic, our interpretation of the floor and ceiling effects models is that we fail to detect design failure. From the estimates in Table 3, furthermore, we have reason to suspect that respondents are willing to reveal vote buying in the list experiment, which could be attributable to sensitivity bias if vote buying is socially desirable.

4.1.2 Uniform error and top-biased error

List experiments are also fallible to various types of measurement error which are common to all surveys, such as error due to respondent inattentiveness, misunderstanding or satisficing (Ahlquist 2018; Blair et al. 2019; Kuhn and Vivyan 2021). Ahlquist (2018) and Blair et al. (2019) focus on two types of measurement error in particular. The first is “uniform error”, defined as a process by which a respondent’s truthful response is replaced by a random uniform draw from the possible answers available, which in turn depends on treatment assignment. The second is “top-biased error”, in which a respondent’s truthful response is randomly replaced with the maximum value available.

These error processes would cause both the DiM estimator and the standard maximum likelihood estimator (MLreg) introduced by Blair and Imai (2012) to be biased (Ahlquist 2018). To investigate the possibility of these sources of design failure we can therefore compare the DiM and MLreg estimators to two alternative ML estimators, introduced by Blair et al. (2019), which model error processes directly (MLreg-top and MLreg-uniform). These ML models are not only robust towards the error process, but also estimate the proportion of respondents which exhibit the error process. In addition, we compare these ML estimators to an unconstrained estimator (ML-unconstrained), introduced by Imai (2011), which relaxes the assumption that parameters for the control items are equal for those with and without the sensitive trait thereby allowing us to investigate whether having experienced vote buying affects the answers to the control items.

¹² This is the same logic as the floor effect in situation iii) when the sensitive behavior is the affirmative answer as both of these situations will allow the respondent to signal the socially desirable behavior.

¹³ We develop this point further in section B.3 of the online appendix.

¹⁴ We report the full results of the floor and ceiling models, as well as Likelihood-ratio tests comparing these models in the online appendix, section B.4.

If the list experiment estimates remain consistent across these models, we can infer that these error processes are not occurring to a substantial extent.¹⁵ We also compare the fit of the models to the data using the AIC, and test whether the models fit the data equally well using likelihood ratio tests (see online appendix B.4).

Figure 2 highlights the diversity of the estimated proportion of individuals who reported vote buying depending on the estimation technique. While the DiM and NLS estimates are almost identical (29%), which is expected as NLS is a generalization of DiM, the standard MLreg yields a much higher estimate (41%). Under the assumption that either the top-biased or uniform error mechanisms are at play, the estimated proportion of individuals drops to just above 20%. Turning to the unconstrained model, the estimated proportion is only around 15%. Since the DiM and NLS estimators tend to be biased in the presence of measurement error, the large discrepancy comparing the simple DiM to the standard MLreg model suggests that the DiM may in fact be overestimating the proportion of individuals who have participated in vote buying. That is, respondents seem to be over-reporting vote buying in the list experiment, as well as in the direct measure. Additionally, the discrepancy across the standard MLreg and the two error models suggests that uniform error and top-biased error account for the overestimation.

Examining the results more closely, several interesting findings come to light. First, we can see that the top-biased error model estimates that 8.6% (95% CI 7.3–10.1%) of the respondents are ‘top-coders’. This is a substantively large proportion, given that prior research considers 3% problematic (Ahlquist 2018). Our interpretation of this result again depends on our assumptions of polarity. Blair et al. (2019) argue that top-biased error should be uncommon, since choosing the maximum response is equivalent to forfeiting confidentiality and revealing a socially undesirable behavior. Under the assumption that choosing the maximum response reveals an undesirable behavior, top-biased error would suggest design failure in the form of respondent inattentiveness, misunderstanding or a technical error in the administration or coding of the survey. However, as we have noted in our discussion of floor- and ceiling-liars, choosing the maximum response may be strategic under the assumption that doing so reveals a desirable behavior.

Second, in the uniform error model, we find that the error is solely located in the treatment group. The model estimates that none of the respondents in the control group exhibit uniform error, while 27.5% (95% CI 21.0–35.0%) in the treated group exhibit uniform error. The presence of uniform error is usually interpreted as evidence of nonstrategic error, resulting for example from respondents providing random answers in order to “satisfice” (Blair et al. 2019, 17). Yet, the fact that uniform error only occurs in the treatment group in our survey suggests that respondents may be strategically choosing a random response as a way of concealing their true sensitive response. In other words, evidence of uniform error could indicate sensitivity bias, rather than design failure.

Third, the unconstrained ML model indicates that there is a strong estimated effect of having reported vote buying in the list experiment, and answering affirmatively to control items. This is evident when looking closer at the models where the intercept for the control items in the ML-unconstrained model is substantively larger for the respondents who

¹⁵ Another explanation for discrepancies across alternative models is that the sensitive topic is rare. Blair et al. (2019) show that in cases where the sensitive topic occurs very infrequently, existing models will return inconsistent results. Since vote buying is widely considered to be prevalent in the 2019 elections in Nigeria (Onuoha and Ojo 2018; Olaniyan 2020; Obe 2019) we do not consider this a likely alternative explanation.

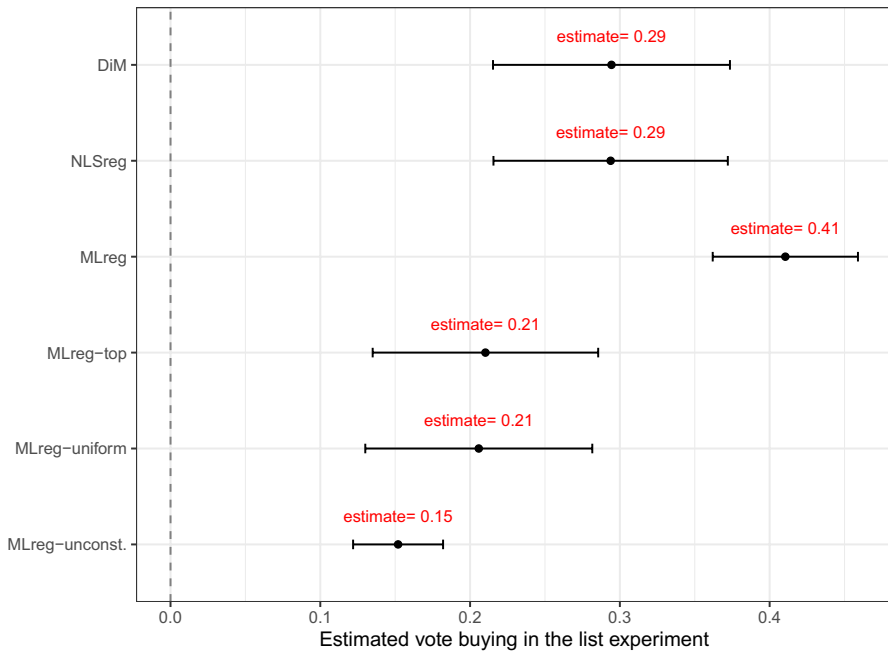


Fig. 2 Estimated proportion of respondents who reported vote buying in the list experiment, using alternative models. This figure shows the estimated proportion of respondents who reported vote buying in the list experiment, using the following estimators: the difference-in-mean (DiM), the nonlinear least squares regression (NLSreg) estimator, the standard maximum likelihood (MLreg) estimator, the top-biased error model (MLreg-top), uniform error models (MLreg-uniform) and the unconstrained maximum likelihood estimator (MLreg-unconst). All models are estimated via the `ictpreg` function in the `List` package for R. With the exception of DiM, all models include age, gender, education, employment and registered voter as covariates. Plotted points represent the mean estimated proportion. Horizontal bars are 90% confidence intervals

reported vote buying compared to the intercept for those who did not report buying. In other words, a large proportion of the ‘top-coders’ identified in the top-biased error model may in fact be individuals who participated in vote buying. This gives additional support to the interpretation of top-coding as evidence of strategic misreporting. Assuming that answering “no” is the sensitive response, choosing the maximum response may be especially preferable for respondents who truly have participated in vote buying: in order for the social benefits of vote buying to be conferred, respondents must be identifiable. So, if vote buying is socially desirable, the presence of top-biased error among those who have participated in vote buying could be viewed as evidence of sensitivity bias.

4.2 Subgroup analysis of sensitivity bias polarity

So far, we have focused on the assumption that the polarity of sensitivity bias is *uni-directional* in the case of vote buying: the assumption that vote buying should be under-reported. In this section we briefly consider a related assumption: that polarity is *uniform* across respondent sub-groups. In contrast to this assumption, prior research suggests that

whether a topic is stigmatised or socially desirable likely varies according to individual characteristics such as political ideology, religious affiliation, gender or race (Aronow et al. (2015, p. 58), Höglinger and Jann (2018, p. 3), Lax et al. (2016, p. 523)). In the case where a large subgroup has reasons to over-report, this could manifest in over-reporting in the aggregate, although the nature of sensitivity bias may push in both directions. Similarly, if the sensitivity bias is in opposite directions for certain sub-groups, then even highly sensitive topics may seem non-sensitive in the aggregate as the over- and under-reporting across subgroups may cancel each other out. Non-uniform polarity, in short, could be an alternative explanation for unexpected list experiment results.

There are theoretical reasons to believe some subgroups of the Nigerian population over-report vote buying, while others under-report vote buying. Specifically, existing literature on the prevalence of vote buying suggest that socio-economic factors, such as income levels and education might enhance the social stigma associated with participation in vote buying (e.g., Gonzalez Ocantos et al. 2014). The close association between ethnic group affiliation and vote choice in Nigeria and the importance of ethnic ties for clientelistic mobilization suggest that social norms surrounding the acceptability of vote buying could differ across groups, not merely (or even primarily) across individuals. Nigeria has vast disparities in the distribution wealth and education across ethnic groups and regions (Rustad and Oestby 2017; Archibong 2018). We could thus expect the relatively more educated and well-off groups in the South and Southwest, such as the Yoruba and Igbo, to be more sensitive to reporting vote buying when asked in a direct question as it violates social norms within more educated and well-off ethnic communities.

In online appendix B.6 we investigate the differences in estimated sensitivity bias across a range of respondent subgroups. It should be noted that the experiment is neither designed nor powered to detect subgroup differences in polarity, and we should be aware of the multiple comparisons problem involved in conducting so many subgroup tests. With these caveats in mind, the analysis does suggest that there may be different polarities across sub-groups. In particular, the Yoruba ethnic group seems to have sensitivity bias in the opposite direction of the non-Yoruba respondents. The result also holds for the Southwest region, where the vast majority (94.8%) of respondents are Yoruba. Perhaps even more interesting is when we look at the results from the Lagos region. Here, the aggregate difference between the direct and list experiment measures are small, indicating that perhaps vote buying may not be sensitive in the region. However, when looking at the different ethnic groups within Lagos we see that the point estimate for Yorubas in Lagos is negative while the point estimates for other ethnicities are positive. These differences are not statistically significant (there are relatively few respondents in the Lagos region), but still serve as a useful illustration that the differences in subgroup sensitivity bias polarity may mask the sensitivity of vote buying.

In addition to differences in sensitivity bias polarity across ethnic groups, we also looked for differences according to whom the respondent thought sponsored the survey. In many contexts, prior research suggests that responses to a sensitive question depend on a respondent's social referent (Blair et al. 2020), or on beliefs about 'who is asking' (Isani and Schlipphak 2022). Keeping in mind the caveats surrounding subgroup analysis, we do find further evidence of non-uniformity. In particular, the small number of respondents who believe that a media actor is the survey sponsor seem to under-report vote buying. There is also a tendency towards under-reporting among the respondents who believed a political actor was the survey sponsor.

That the perceived survey sponsor may impact sensitivity bias polarity further highlights the issues of survey sponsor misconceptions detected by previous research (Isani and Schlipphak 2022). It also supports our earlier findings which suggest that the misreporting we observe is strategic, rather than design failure. Indeed, a core assumption of the “social reference theory” of sensitivity bias is that individuals adjust their responses according to who they believe can access the data and according to the likely consequences (positive or negative) associated with their response (Blair et al. 2020, p. 1299). It follows that one observable indicator of sensitivity bias –as opposed to inattentiveness– is that misreporting should vary by perceived survey sponsor.

5 Conclusion

Our exploratory analysis aimed to establish whether the over-reporting of vote buying we observed in our post-election survey can be attributed to sensitivity bias, or whether it is simply the result of a failed list experiment. By submitting the survey data to a series of test of design, we found some evidence that over-reporting is strategic in nature. First, when testing for violations of the “no design effects” and “no liars” assumptions of list experiments, we found no evidence indicative of design failure. Tests for “floor” and “ceiling” effects suggest that respondents do not conceal vote buying in the list experiment. On the contrary, respondents appear to intentionally reveal vote buying. Second, when examining the extent of uniform and top-biased error, we find evidence that respondents strategically select random responses, and that it is likely that who reveal vote buying by top-coding have in fact participated in vote buying. Our conclusion is that the over-reporting we observe is consistent with sensitivity bias in a context where vote buying is considered socially desirable, legitimate or beneficial.

The contribution we make is simultaneously methodological and theoretical: we have shown how our theoretical assumptions about what is desirable and what is stigmatised guide our interpretation of list experiment results. While prior research has shown how standard diagnostic tests can fail to accurately distinguish between strategic and non-strategic measurement error in indirect questioning techniques (e.g., Kuhn and Vivyan 2021; Höglinger and Jann 2018), our focus is specifically on the role of theoretical assumptions in the interpretation of these tests. In particular, we show how uni-directional assumptions about polarity guide what is considered evidence of strategic or non-strategic error, and consequently, what is interpreted as sensitivity bias or design failure.

These assumptions are particularly strong because they imply uniformity across respondent subgroups (e.g., men and women have the same opinion on what the sensitive response is) and across respondents’ social referents (e.g., individual choices to conceal or exaggerate do not depend on who is asking). We probed some of these assumptions via subgroup analysis, and found initial evidence of non-uniformity. However, we stopped short of confirmatory analyses of subgroup differences in sensitivity bias, given that our list experiment was not designed nor sufficiently powered to detect these differences. We view this an important direction for future research. Considering a range of controversial political topics– vote choice, prejudice, abortion and same-sex marriage– it is challenging to imagine a setting in which different groups agree on what the socially correct response is. Given that uni-directional assumptions about polarity may be problematic, we encourage future research to probe additional assumptions both in relation to theory, and in relation to the analysis of list experiments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11135-023-01740-6>.

Acknowledgements For valuable feedback, we thank Patrick M. Kuhn, Erik Skoog, Annkatrin Deglow, participants of the 2022 International Studies Association (ISA) Annual Conference, and participants of the 2021 Elections and Violence Workshop at Uppsala University.

Author Contributions All authors contributed to the study conception and empirical approach. The survey design and data collection was prepared by Hanne Fjelde and analyses were performed by David Randahl and Sophia Hatz. All authors contributed to the writing of the manuscript and read and approved the final manuscript.

Funding Open access funding provided by Uppsala University. This work was supported by the Swedish Research Council (grant VR 2016-05833), Knut and Alice Wallenberg Foundation (KAW 2017.0141) and Riksbankens Jubileumsfond (grant RJ M21-0002).

Availability of data and materials All data files necessary to replicate the analysis presented in the article will be made available upon publication at [LINK]

Code availability All scripts necessary to replicate the analysis presented in the article will be made available upon publication at [LINK]. R 3.6.2 was the statistical package used in this study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval We received ethical approval for the survey and the experiment from the Swedish Ethical Review Board on 2018-11-07 (ID no 2018/409)

Consent to participate Informed consent was obtained from all participants in the survey and experiment.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahlquist, J.S.: List experiment design, non-strategic respondent error, and item count technique estimators. *Polit. Anal.* **26**(1), 34–53 (2018). <https://doi.org/10.1017/pan.2017.31>
- Archibong, B.: Historical origins of persistent inequality in Nigeria. *Oxf. Dev. Stud.* **46**(3), 325–347 (2018)
- Aronow, P.M., Coppock, A., Crawford, F.W., Green, D.P.: Combining list experiment and direct question estimates of sensitive behavior prevalence. *J. Surv. Stat. Methodol.* **3**(1), 43–66 (2015)
- Blair, G., Chou, W., Imai, K.: List experiments with measurement error. *Polit. Anal.* **27**(4), 455–480 (2019)
- Blair, G., Coppock, A., Moor, M.: When to worry about sensitivity bias: a social reference theory and evidence from 30 years of list experiments. *Am. Polit. Sci. Rev.* **114**(4), 1297–1315 (2020)
- Blair, G., Imai, K.: Statistical analysis of list experiments. *Polit. Anal.* **20**(1), 47–77 (2012)
- Bratton, M.: Vote buying and violence in Nigerian election campaigns. *Elect. Stud.* **27**(4), 621–632 (2008). <https://doi.org/10.1016/j.electstud.2008.04.013>
- Cammett, M.: *Compassionate Communalism: Welfare and Sectarianism in Lebanon*. Cornell University Press, Ithica and London (2014)

- Carkoglu, A., Aytaç, S.E.: Who gets targeted for vote-buying? evidence from an augmented list experiment in Turkey. *Eur. Polit. Sci. Rev.* **7**(4), 547–566 (2015)
- Castro Cornejo, R., Beltrán, U.: List experiments, political sophistication, and vote buying: experimental evidence from Mexico. *J. Polit. Lat. Am.* **12**(2), 219–234 (2020). <https://doi.org/10.1177/1866802x20937713>
- Castro Cornejo, R., Beltrán, U.: Who receives electoral gifts? It depends on question wording: experimental evidence from Mexico. *Polit. Behav.* **44**(1), 227–255 (2022). <https://doi.org/10.1007/s11109-020-09618-1>
- Chabal, P., Daloz, J.P.: *Africa Works: Disorder as Political Instrument*. International African Institute in assoc. with James Currey, Indiana University Press, London (1999)
- Corstange, D.: Clientelism in competitive and uncompetitive elections. *Comp. Polit. Stud.* **51**(1), 76–104 (2018)
- EU Election Observation Mission.: *Nigeria 2019 Final Report* (2019)
- Finan, F., Schechter, L.: Vote-buying and reciprocity. *Econometrica* **80**(2), 863–881 (2012). <https://doi.org/10.3982/ecta9035>
- Fisher, R.J.: Social desirability bias and the validity of indirect questioning. *J. Consum. Res.* **20**(2), 303–315 (1993)
- Gadjanova, E.: Electoral clientelism as status affirmation in Africa: Evidence from Ghana. *J. Mod. Afr. Stud.* **55**(4), 593–621 (2017). <https://doi.org/10.1017/S0022278X17000416>
- Gelman, A.: Thinking of Doing a List Experiment? Here’s a List of Reasons You Should Think Again (2014). Accessed 24 Aug 2021
- Glynn, A.N.: What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opin. Q.* **77**(S1), 159–172 (2013)
- Gonzalez-Ocantos, E., De Jonge, C.K., Meléndez, C., Osorio, J., Nickerson, D.W.: Vote buying and social desirability bias: experimental evidence from Nicaragua. *Am. J. Polit. Sci.* **56**(1), 202–217 (2012)
- Gonzalez Ocantos, E., de Jonge, C.K., Nickerson, D.W.: The conditionality of vote-buying norms: experimental evidence from Latin America. *Am. J. Polit. Sci.* **58**(1), 197–211 (2014). <https://doi.org/10.1111/ajps.12047>
- Höglinger, M., Jann, B.: More is not always better: an experimental individual-level validation of the randomized response technique and the crosswise model. *PLoS One* **13**(8), 1–22 (2018). <https://doi.org/10.1371/journal.pone.0201770>
- Imai, K.: Multivariate regression analysis for the item count technique. *J. Am. Stat. Assoc.* **106**(494), 407–416 (2011). <https://doi.org/10.1198/jasa.2011.ap10415>
- Isani, M., Schlipphak, B.: Who is asking? The effect of survey sponsor misperception on political trust: evidence from the afrobarometer. *Qual. Quant.* (2022). <https://doi.org/10.1007/s11135-022-01517-3>
- Kerr, N.: Vote selling in the 2015 Nigeria elections: Challenges of using list experiments to gauge social desirability bias (2018)
- Kiewiet De Jonge, C.P.: Who lies about electoral gifts? Experimental evidence from Latin America. *Public Opin. Q.* **79**(3), 710–739 (2015). <https://doi.org/10.1093/poq/nfv024>
- Konishi, S., Kitagawa, G.: *Information Criteria and Statistical Modeling*. Springer, New York (2008)
- Kramon, E.: Where is vote buying effective? Evidence from a list experiment in Kenya. *Elect. Stud.* **44**, 397–408 (2016)
- Kramon, E., Weghorst, K.: (Mis) measuring sensitive attitudes with the list experiment: solutions to list experiment breakdown in Kenya. *Public Opin. Q.* **83**(S1), 236–263 (2019)
- Krumpal, I.: Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quant.* **47**(4), 2025–2047 (2013)
- Kuhn, P.M., Vivyan, N.: The misreporting trade-off between list experiments and direct questions in practice: partition validation evidence from two countries. *Polit. Anal.* **1**, 22 (2021). <https://doi.org/10.1017/pan.2021.10>
- Lawson, C., Greene, K.F.: Making clientelism work: how norms of reciprocity increase voter compliance. *Comp. Polit.* **47**(1), 61–85 (2014)
- Lax, J.R., Phillips, J.H., Stollwerk, A.F.: Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. *Public Opin. Q.* **80**, 510–533 (2016)
- Mares, I., Muntean, A., Petrova, T.: Economic intimidation in contemporary elections: evidence from Romania and Bulgaria. *Gov. Oppos.* **53**(3), 486 (2018)
- Mares, I., Young, L.E.: *Conditionality & Coercion: Electoral Clientelism in Eastern Europe*. Oxford University Press, Oxford (2019)
- Nugent, P.: Banknotes and symbolic capital: Ghana’s elections under the fourth republic. In: Basedau, M., Erdmann, G., Mehler, A. (eds.) *Votes, Money and Violence: Political Parties and Elections in*

- Sub-Saharan Africa. Nordic Africa Institute and Scottsville, University of Kwazulu-Natal Press, Uppsala (2007)
- Obe, A.: Aspirations and realities in Africa: Nigeria's emerging two-party system? *J. Democr.* **30**(3), 109–123 (2019). <https://doi.org/10.1353/jod.2019.0046>
- Okakwa, E.: Nigeria's 'Freest' Election Witnessed Vote-buying 'Worth N120m to N1bn' (2019). Accessed 02 Dec 2021
- Olaniyan, A.: Election sophistication and the changing contours of vote buying in Nigeria's 2019 general elections. *Round Table* **109**(4), 386–395 (2020)
- Onuoha, F., Ojo, J.: Practice and Perils of Vote Buying in Nigeria's Recent Elections (2018). Accessed 02 Dec 2021
- Riambau, G., Ostwald, K.: Placebo statements in list experiments: evidence from a face-to-face survey in Singapore. *Polit. Sci. Res. Methods* **9**(1), 172–179 (2021). <https://doi.org/10.1017/psrm.2020.18>
- Rinken, S., del Amo, S.P., Rueda, M., Cobo, B.: No magic bullet: estimating anti-immigrant sentiment and social desirability bias with the item-count technique. *Qual. Quant.* **55**, 2139–2159 (2021)
- Rustad, S., Oestby, G.: Education and systematic group inequalities in Nigeria. *Prio Conflict Trends* **3**, 1–4 (2017)
- Sakariyau, R.T., Aliu, F.L., Adamu, M., et al.: The phenomenon of money politics and Nigeria's democratization: an exploration of the fourth Republic. *J. Soc. Econ. Res.* **2**(1), 1–9 (2015)
- Schaffer, F.C. (ed.): *Elections for Sale: the Causes and Consequences of Vote Buying*. Lynne Rienner Publishers, Boulder (2007)
- Scott, J.C.: Patron-client politics and political change in Southeast Asia. *Am. Polit. Sci. Rev.* **66**(01), 91–113 (1972)
- Stokes, S.: Perverse accountability: a formal model of machine politics with evidence from Argentina. *Am. Polit. Sci. Rev.* **99**(3), 315–325 (2005)
- Stokes, S.: Is vote buying undemocratic? In: Schaffer, F.C. (ed.) *Elections for Sale: The Causes and Consequences of Vote Buying*, pp. 81–100. Lynne Rienner, Boulder (2007)
- Tourangeau, R., Yan, T.: Sensitive questions in surveys. *Psychol. Bull.* **133**(5), 859–883 (2007). <https://doi.org/10.1037/0033-2909.133.5.859>
- Tsuchiya, T., Hirai, Y., Ono, S.: A study of the properties of the item count technique. *Public Opin. Q.* **71**(2), 253–72 (2007)
- van de Walle, N.: Meet the new boss, same as the old boss? In: Kitschelt, H., Wilkinson, S.I. (eds.) *Patrons, Clients and Policies*. Cambridge University Press, Cambridge (2007)

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