Supplementary Materials for:

When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments

Graeme Blair[†] Alexander Coppock[‡] Margaret Moor[§]

Contents

[†]Graeme Blair is Assistant Professor of Political Science, University of California, Los Angeles. `https://graemeblair.com`

[‡]Alexander Coppock is Assistant Professor of Political Science, Yale University. `https://alexandercoppock.com`

[§]Margaret Moor is a graduate of Yale College, Class of 2018.

A. Perceptions of Survey Sponsors Vary by Individual and Context

In the Afrobarometer face-to-face survey conducted in several countries in Sub-Saharan Africa, the last question in every survey (since the second Afrobarometer round) asks who respondents think are responsible for the survey. The question text is, "`Just one more question: Who do you think sent us to do this interview?`" Responses are coded by Afrobarometer from recorded verbatim responses. In Figure A.1, we display the proportion of responses to each answer option overall (left panel) and the proportion responding that the "government" is responsible for the survey across countries (right panel). The figure shows that responses vary substantially across respondents, and across countries. Impression management concerns and the perceived risks of disclosure are likely heterogeneous across respondents and contexts.

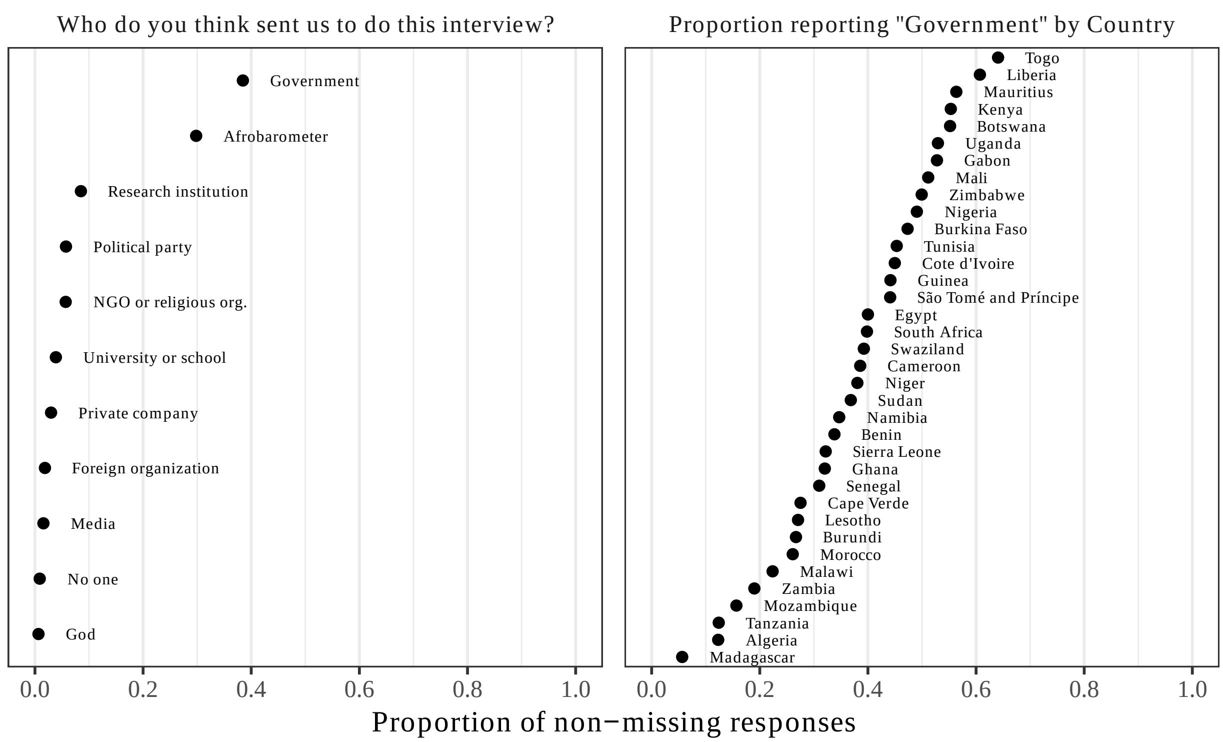


Figure A.1: Beliefs about the organization who sent the interviewer vary both across individuals (left panel) and across countries (right panel). Data from Afrobarometer Round 6 (2014–2015). The question text is "Just one more question: Who do you think sent us to do this interview?" and responses were coded by Afrobarometer from recorded verbatim responses.

B. How the List Experiment Addresses Sensitivity Bias

The list experiment obscures individual responses to the sensitive item, but still allows analysts to estimate sample quantities including sensitive item prevalence and other relevant quantities. With a set of N individuals indexed by i, we randomly assign each to a treatment group ($T_i = 1$) or a control group ($T_i = 0$). In the control group, we ask respondents for a count of the number of "yes" responses to J control items indexed by j. In the treatment group, we ask respondents for a count of the number of "yes" responses to a set of $J + 1$ items, the J control items plus the sensitive item. We define two sets of potential outcomes: $Z_{ij}(t)$ for $t = 0, 1$. The observed outcome is defined as $Y_i = Y_i(T_i)$.

A main aim of researchers is to estimate the sample prevalence of the sensitive item: $\pi^* = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$. In order to identify this quantity, four assumptions must be invoked. These are described in Imai (2011), but we recapitulate them here. First, we need the standard assumptions for identifying the average treatment effect in an experiment: noninterference and the ignorability of the treatment status. Noninterference requires that subjects' outcomes depend only on their own treatment status and not on that of other subjects. In list experiments, noninterference is typically assured by design because subjects take the surveys separately. Ignorability requires that the treatment be independent of the potential outcomes $Y_i(1)$ and $Y_i(0)$ and is guaranteed by design in list experiments because the treatment is randomized.

Two additional assumptions are required in order to interpret this treatment effect as the prevalence rate of the sensitive item. No design effects assumes that responses to the control items do not differ in treatment and control. This assumption would be violated if the presence of the sensitive item changes how subjects respond to the control items. Formally, the no design effects assumption states that for all respondents i, $\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1)$. No liars assumes that respondents do not misreport the "yes" or "no" response to the sensitive item. The no liars assumption states that for all respondents i, $Z_{i,J+1}(1) = D_i^*$. Substantively, no liars means that list experiment responses are not distorted by sensitivity bias. The protection provided by the list experiment removes the threat of costs because the social referent cannot learn subjects' responses.

No liars might be violated if treatment group subjects' true response to the list experiment would be "all" or "none," but they report a different value instead. An answer of "none" would identify them as a "no" to the sensitive item and an answer of "all" would identify them as a "yes" to the sensitive item. For these respondents, the list experiment offers no protection from the aggregation with the control items, so we should not expect a change in the self-presentation pressures or the risk of disclosure. Glynn (2013) describes this specific violation of no liars as floor and ceiling effects. No liars would also be violated if subjects were unable to admit the truth to themselves.

Violations of no design effects occur when respondents evaluate the control items differently in treatment and control. Respondents may be affected simply by the number of items in a list, so in the treatment group which has one more item than control respondents may change responses to the control items (Flavin and Keane 2009). If respondents evaluate items in a list relative to each other, the addition of a new item may change their evaluations of the control items. Indeed, even if respondents do not evaluate items relative to one another, the addition of the sensitive item may simply act as a frame that changes how they think about other items.

Design effects may also be induced by the presence of the sensitive item in the treatment group list due to its sensitivity. Scholars worry that adding the sensitive item triggers impression management concerns generally, and that may affect responses to the control items. Zigerell (2011) notes that respondents may want to send a strong signal that they do are not answering the sensitive item in the affirmative by deflating their responses to the control items to be closer to or at a zero response.

Under noninterference, ignorability, no design effects, and no liars, the sample sensitive item prevalence is nonparametrically identified. We estimate this quantity using the difference-in-means estimator, which is an unbiased estimator under these assumptions.[1]

Other quantities beyond the sensitive item prevalence have been of interest to political scientists. Subgroup prevalence (analogous to conditional average treatment effects in standard experimental settings) and their differences can be estimated with the same tools and justifications. For surveys that also include a direct question on the same topic (such as the Kenya postelection survey reported in Kramon 2016), the difference between the list experiment and the direct question is estimate of sensitivity bias (Janus 2010; Blair and Imai 2012).

[1] The difference-in-means estimator is not the only way to estimate the prevalence rate. Other estimators, such as the nonlinear least squares and maximum likelihood procedures whose main purpose is the estimation of multiple regression coefficients, may generate more precise estimates of the prevalence rate, but do so at the cost of additional modeling assumptions (Imai 2011; Blair et al. 2019).

C. Variance Derivations

Variance of the direct question estimator

In the main text, we use the following expression to describe the variance of the direct question estimator of in terms of the sample size n, the true prevalence rate π^*, and the level of sensitivity bias δ:

$$\mathbb{V}(\widehat{\pi}) = \frac{\pi^*(1-\pi^*) + \delta(1-\delta) - 2(\delta - \pi^*\delta)}{n-1}$$

Subject i's true latent trait is D_i^*. The response that subject is would give to the direct question is D_i. We define the difference between these as $W_i \equiv D_i^* - D_i$. Sensitivity bias, therefore is the expectation of W_i: $\delta = \mathbb{E}[W_i]$ The direct question estimator $\widehat{\pi}$ is the sample mean $\widehat{\pi} = \frac{1}{n}\sum_1^n D_i$, which has variance $\frac{\mathbb{V}(D_i)}{n-1}$ by standard formulas. Since $D_i = D_i^* - W_i$, the variance of D_i can be written $\mathbb{V}(D_i^*) + \mathbb{V}(W_i) - 2\text{cov}(D_i^*, W_i)$. We need an expression for $\text{cov}(D_i^*, W_i)$. Here we add an additional assumption of monotonicity that states that the value of W_i is either 0 or 1 for all subjects, as in the typical underreporting case. An analogous expression holds in the overreporting case. Monotonicity may not hold in the entire subject pool, but it may be possible to construct subgroups for which the monotonicity holds within the subgroup.

$$\begin{aligned}
\text{cov}(D_i^*, W_i) &= \mathbb{E}[(D_i^* - \mathbb{E}[D_i^*])(W_i - \mathbb{E}[W_i])] && (1)\\
&= \mathbb{E}[(D_i^* - \pi^*)(W_i - \delta)] && (2)\\
&= \mathbb{E}[(D_i^* W_i)] - \mathbb{E}[D_i^*\delta] - \mathbb{E}[\pi^* W_i] + \mathbb{E}[\pi^*\delta] && (3)\\
&= \delta - \pi^*\delta - \pi^*\delta + \pi^*\delta && (4)\\
&= \delta - \pi^*\delta && (5)
\end{aligned}$$

Equation (1) holds from the definition of the covariance; (2) relabels the expectation of the sensitive item as π^* and the expectation of the withholding indicator W_i as δ; (3) distributes terms and uses the linearity property of expectations; (4) simplifies using the definitions of the sensitivity bias δ and the sensitive item prevalence π^* and the monotonicity assumption in order to simplify $\mathbb{E}(D_i^* W_i)$ into δ; and (5) combines terms.

Plugging this expression back in, we see that

$$\mathbb{V}(\widehat{\pi}) = \frac{\mathbb{V}(D_i^*) + \mathbb{V}(W_i) - 2\text{cov}(D_i^*, W_i)}{n-1} \tag{6}$$

$$= \frac{\pi^*(1-\pi^*) + \delta(1-\delta) - 2(\delta - \pi^*\delta)}{n-1} \tag{7}$$

We invoke the monotonicity assumption in order to be able to express this variance of direct question responses in terms of the sensitive item proportion (π^*) and sensitivity bias (δ).

Variance of the list experiment estimator

In the main text, we use the following expression to describe the variance of the list experiment estimator ($\widehat{\pi^*}$) under a balanced design (i.e., $m = N/2$) in terms of the sample size N, the true prevalence rate π^*, the variance of the control item response $\mathbb{V}(Y_i(0))$, and the covariance of the control item response with the sensitive item $\text{cov}(Y_i(0), D_i^*)$.

$$\mathbb{V}(\widehat{\pi^*}) = \frac{1}{N-1}\left\{\pi^*(1-\pi^*) + 4\mathbb{V}(Y_i(0)) + 4\text{cov}(Y_i(0), D_i^*)\right\}$$

Here we derive that expression for designs that may or may not be balanced. Equation 1 begins with the square of Eq. 3.4 in Gerber and Green (2012), which defines the variance of the difference-in-means estimator under complete random assignment as follows.

$$\mathbb{V}(\widehat{\pi^*}) = \frac{1}{N-1}\left\{\frac{m}{N-m}\mathbb{V}(Y_i(0)) + \frac{N-m}{m}\mathbb{V}(Y_i(1)) + 2\text{cov}(Y_i(0), Y_i(1))\right\} \tag{8}$$

$$= \frac{1}{N-1}\left\{\frac{m}{N-m}\mathbb{V}(Y_i(0)) + \frac{N-m}{m}\mathbb{V}(Y_i(0) + D_i^*) + 2\text{cov}(Y_i(0), Y_i(0) + D_i^*)\right\} \tag{9}$$

$$\begin{aligned} = \frac{1}{N-1}\Bigg\{ & \frac{m}{N-m}\mathbb{V}(Y_i(0)) + \\ & \frac{N-m}{m}\Big(\mathbb{V}(Y_i(0)) + \mathbb{V}(D_i^*) + 2\text{cov}(Y_i(0), D_i^*)\Big) + \\ & 2\{\text{cov}(Y_i(0), D_i^*) + \mathbb{V}(Y_i(0))\}\Bigg\} \end{aligned} \tag{10}$$

$$= \frac{1}{N-1}\left\{\frac{N-m}{m}\mathbb{V}(D_i^*) + \left(\frac{m}{N-m} + \frac{N-m}{m} + 2\right)\mathbb{V}(Y_i(0)) + 2\left(\frac{N-m}{m} + 1\right)\text{cov}(Y_i(0), D_i^*)\right\} \tag{11}$$

$$= \frac{1}{N-1}\left\{\frac{N-m}{m}\pi^*(1-\pi^*) + \left(\frac{m}{N-m} + \frac{N-m}{m} + 2\right)\mathbb{V}(Y_i(0)) + 2\left(\frac{N-m}{m} + 1\right)\text{cov}(Y_i(0), D_i^*)\right\} \tag{12}$$

In equation (9), we assume no liars and no design effects (Imai 2011), so $Y_i(1) = Y_i(0) + D_i^*$. Equation (10) follows from the definitions of variance and covariance. Equation (11) collects terms. Equation (12) reexpresses the variance of the sensitive item $\mathbb{V}(D_i^*)$ in terms of the true prevalence rate π^*.

The equation in the main text is a simplified version of Equation (12) under a balanced design ($m = N/2$), which allows us to simplify the expression considerably.

D. Empirical Distributions of Simulation Parameters

In Figure D.2, we present the empirical distributions from our meta analysis data of the four parameters used in our design tradeoff simulations. The means from each empirical distribution (black lines) are used as the simulation parameter. The statistics are calculated from the maximum subset of the data for which they are available.

Importantly, due to the fundamental problem of causal inference we are unable to directly calculate the covariance between the control item count and the true sensitive item response, $\text{cov}(Y_i(0), D_i^*)$. Instead, we are calculate $\text{cov}(Y_i(0), D_i)$, the covariance between the control item count and the response to the direct question. These covariances are quite small, possibly reflecting the success of list experiment designers in following the design advice of Glynn (2013) to choose negatively correlated control items. If control items were perfectly negatively correlated, the control item count would take a constant value for all subjects and the covariance with the sensitive trait would be exactly zero.

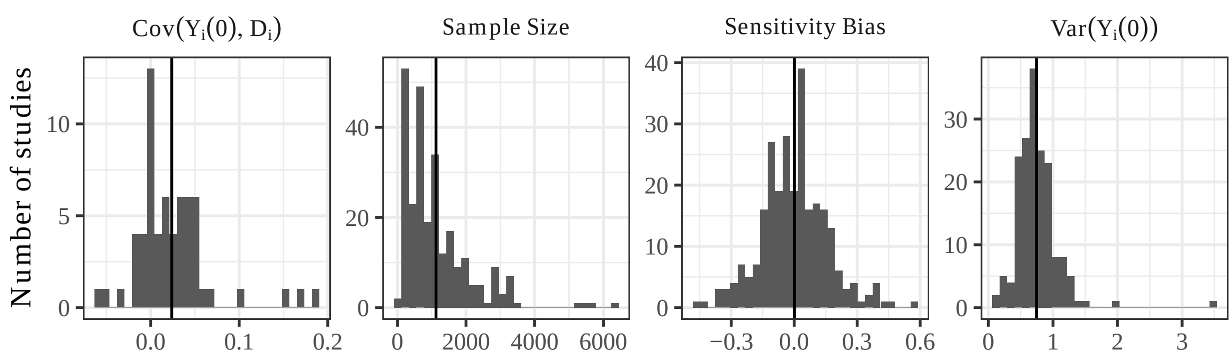


Figure D.2: Empirical Distribution of Each Parameter Used in Design Simulations.

E. Sensitivity Bias by Research Area

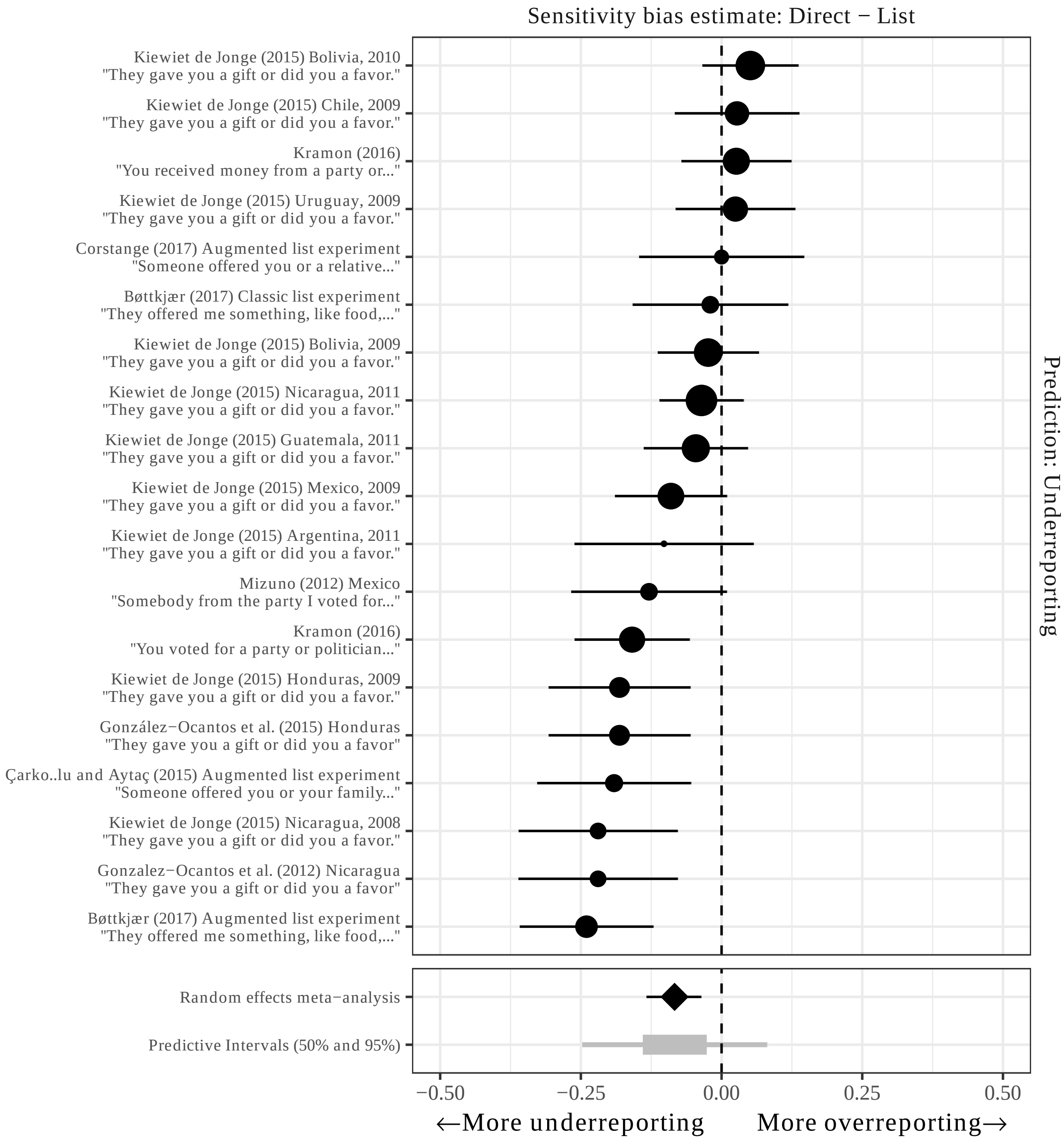


Figure E.3: Estimates of Sensitivity Bias for Vote Buying

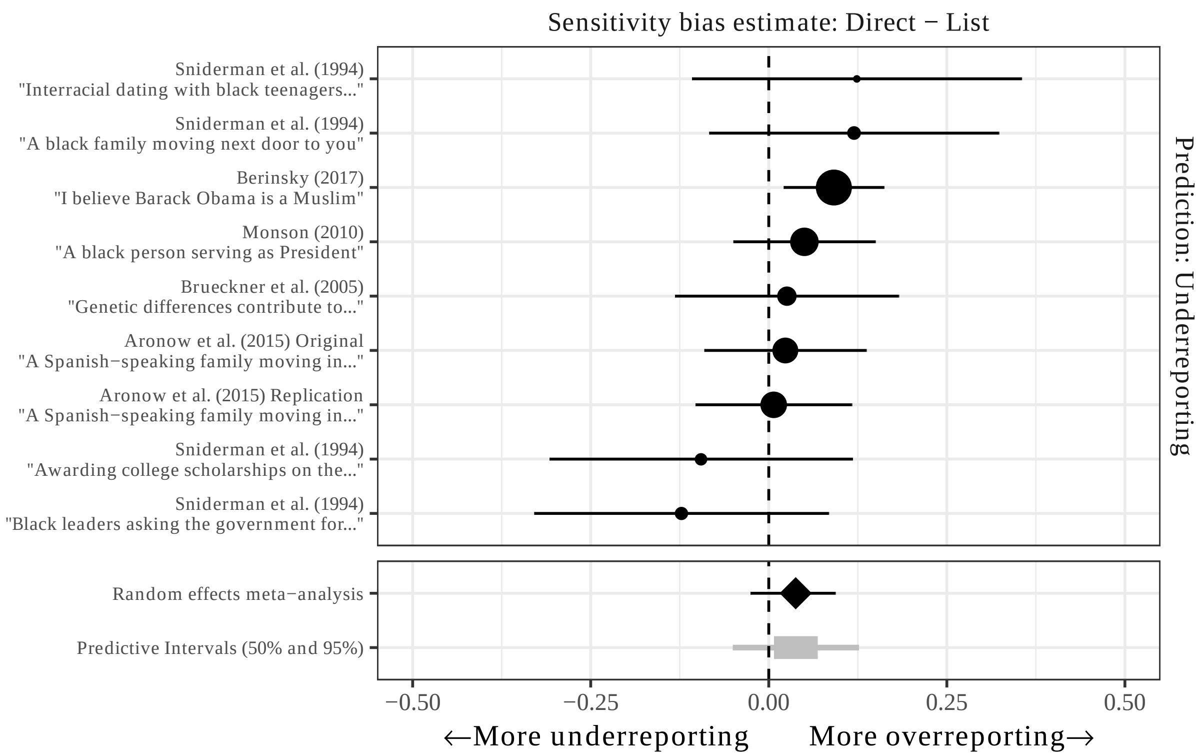
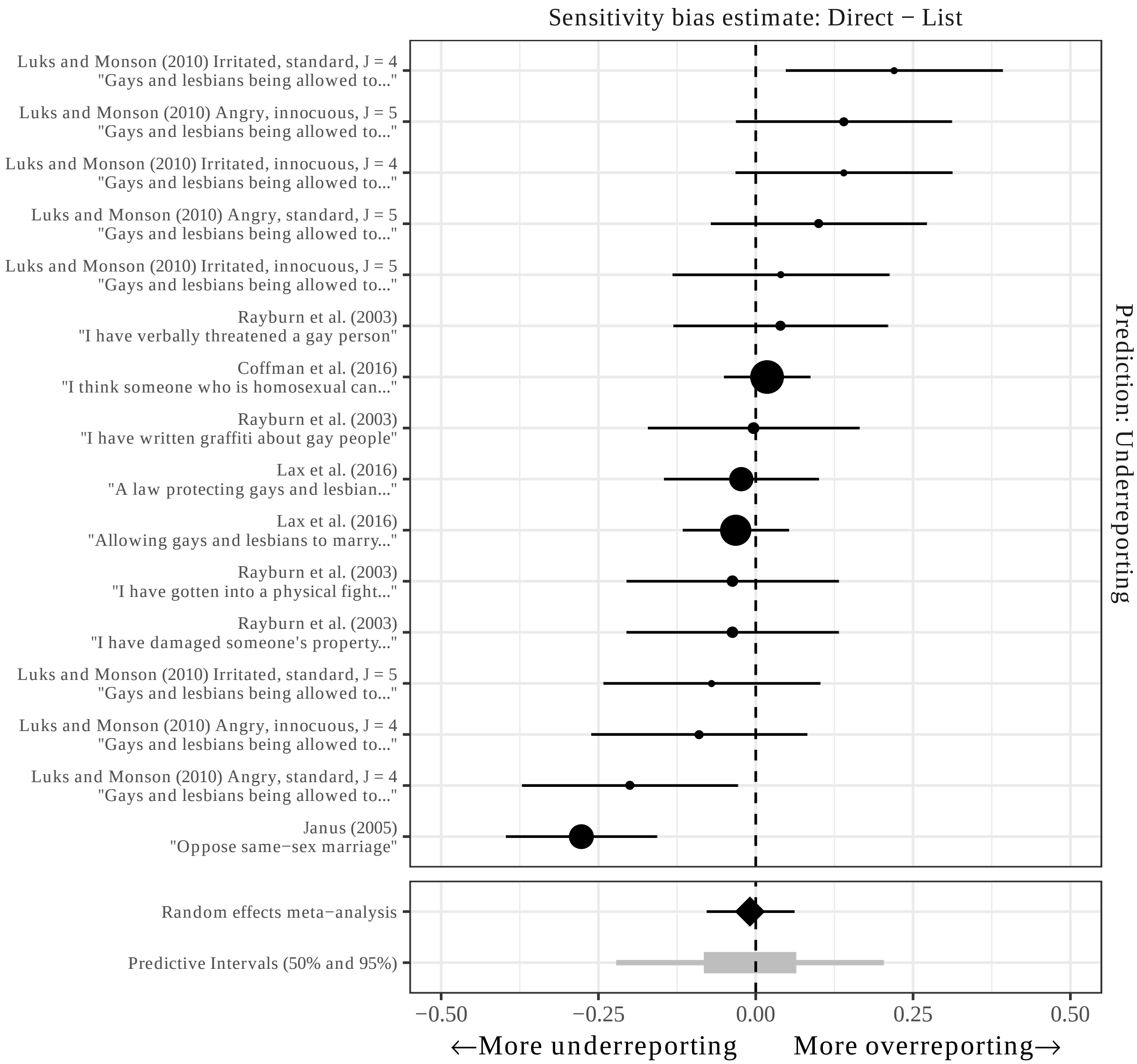
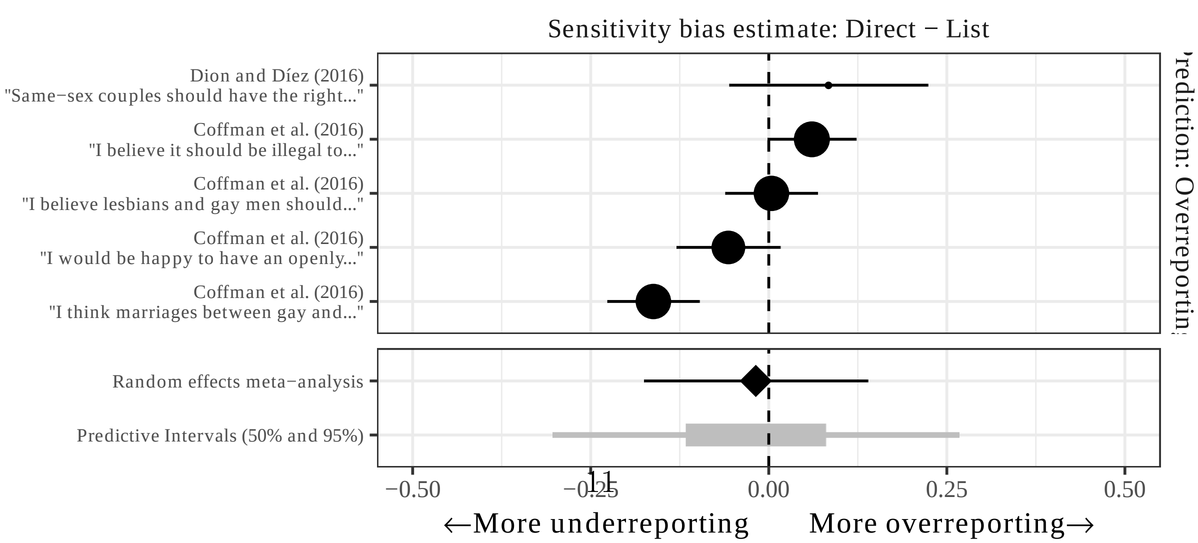


Figure E.4: Estimates of Sensitivity Bias for Racial Prejudice



(a) Prediction: Underreporting



(b) Prediction: Overreporting

Figure E.5: Estimates of Sensitivity Bias for Sexual Orientation Prejudice

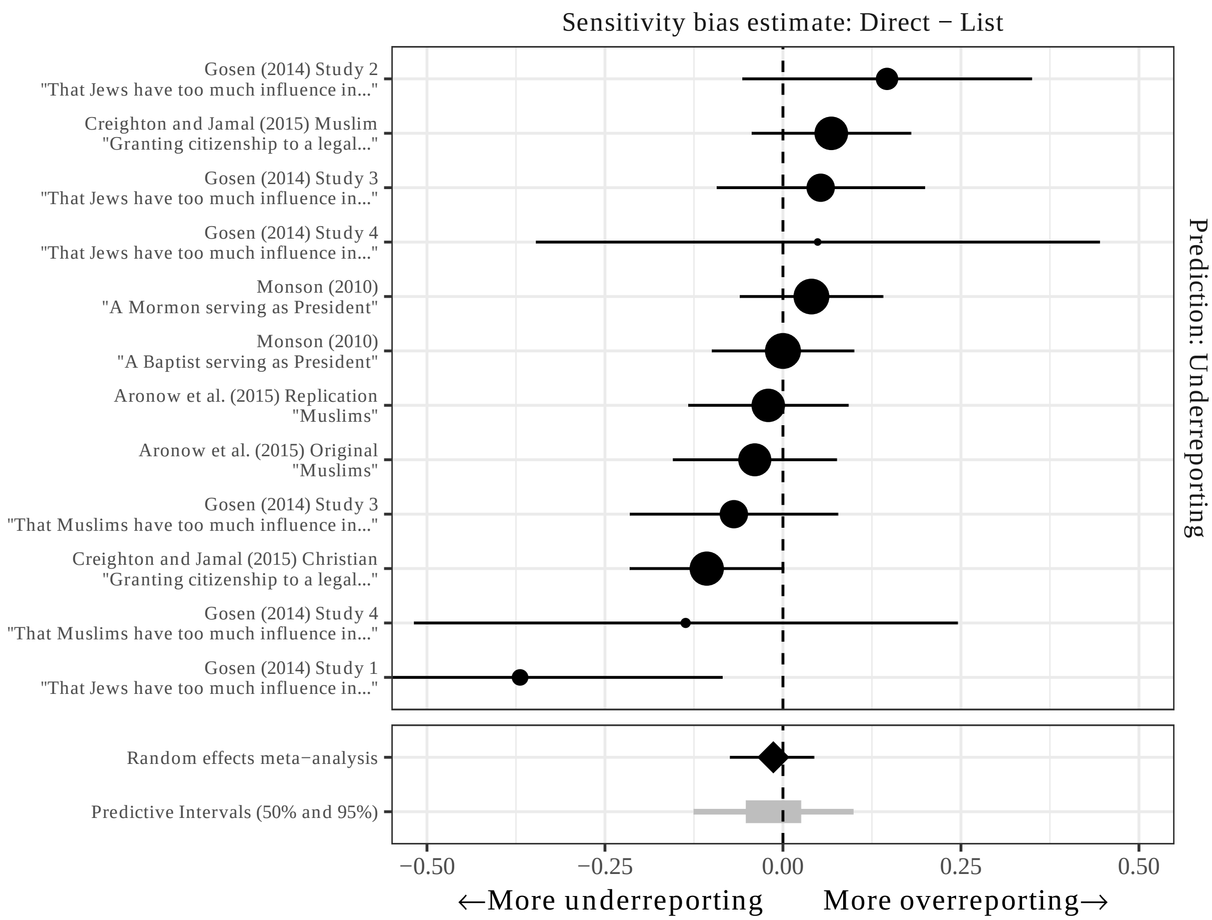


Figure E.6: Estimates of Sensitivity Bias for Religious Prejudice

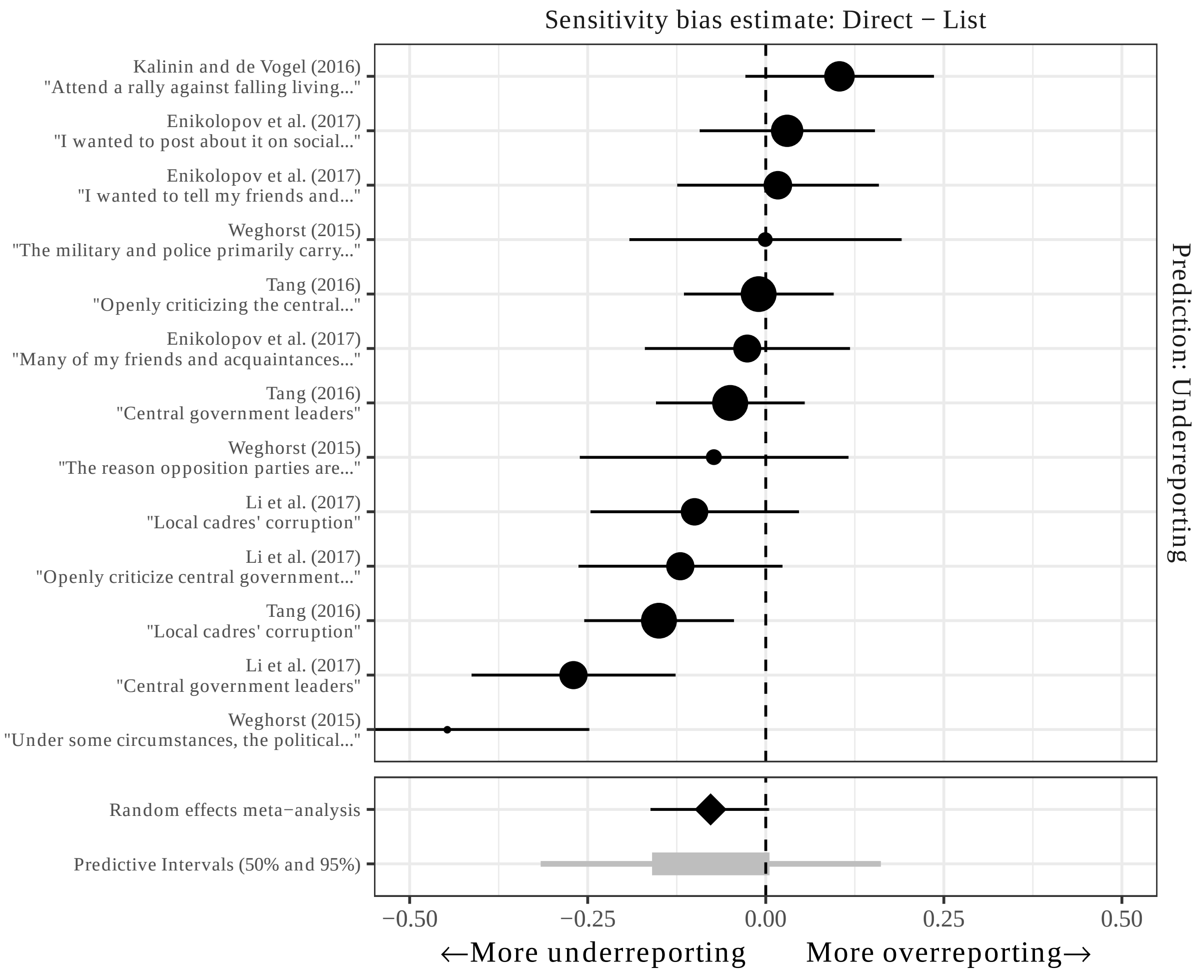


Figure E.7: Estimates of Sensitivity Bias for Political Attitudes in Authoritarian Regimes

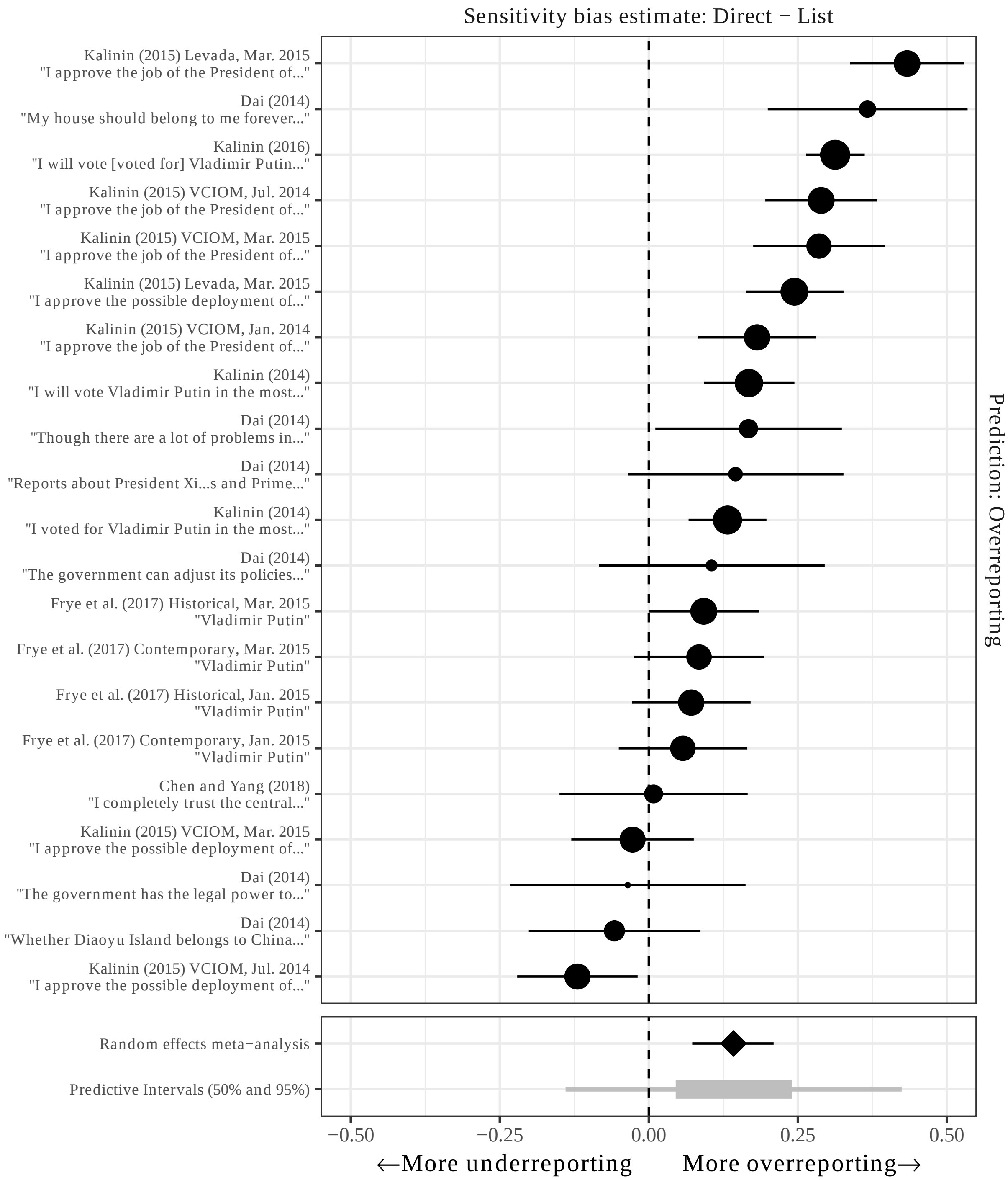


Figure E.8: Estimates of Sensitivity Bias for Political Attitudes in Authoritarian Regimes

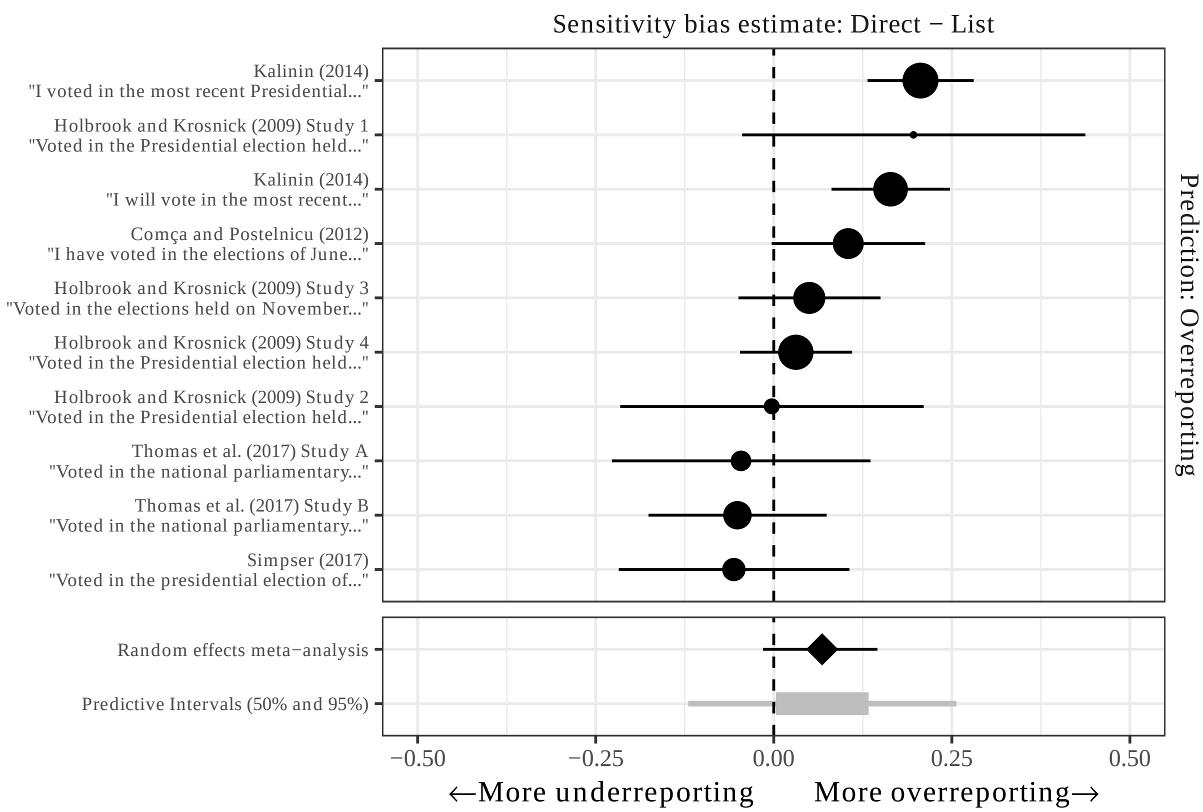


Figure E.9: Estimates of Sensitivity Bias for Turnout

F. Sensitivity Bias by Predicted Direction of Misreporting

In the main text, we zoomed in on results from four political science literatures. In this section, we zoom out to the full set of studies for which we have both list and direct estimates, regardless of discipline or topic. Figure F.10 plots the direct question estimate against the list experiment estimate separately by the predicted direction of sensitivity bias. The point size is proportional to the standard error of the difference estimate (more precise estimates are larger). We also present 95% confidence intervals for both estimates. The regression line overlaid on top of the raw estimates is fit via Deming regression (Deming 1943), an errors-in-variables model, which is appropriate given the measurement error in both the left-hand and right-hand sides of the equation. We estimate measurement error with the standard errors of the direct and list estimates.

First, we see that the direct and list estimates are highly correlated – prima facie evidence that whatever the measurement properties of direct questions and list experimentation, they appear to measure the same latent quantity. One measure of the strength of this correlation is the slope of the Deming regressions, both of which are close to 1. Second, as shown in Table 4, the average sensitivity bias in the case of underreporting is -3 points (SE: 1 point). For overreporting, the bias is much larger at +12 points (SE: 2 points). This asymmetry can be observed by comparing the two panels of Figure F.10. For overreporting, points lie overwhelmingly above the 45 degree line, whereas for underreporting points cluster tightly around it.

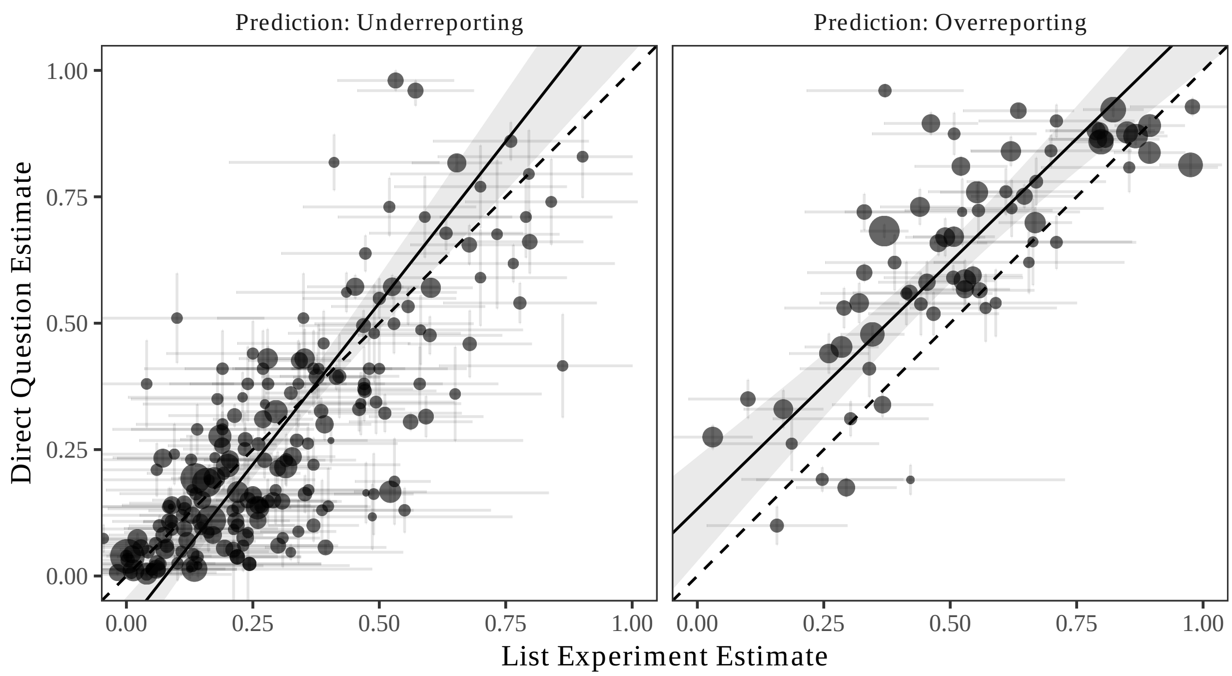


Figure F.10: List Experiment Estimates are Correlated with Direct Question Estimates, and Across Domains There is Sensitivity Bias Especially when Overreporting is Predicted. Estimates of the prevalence rate of the sensitive item from the list experiment (x axis) and from a direct question (y axis) are presented as points along with 95% confidence intervals of each estimate (light gray lines) with point size proportional to the weight from a Deming errors-in-variables regression. The Deming regression model fit (solid line) is presented along with its 95% confidence interval (gray area). The 45% degree line, representing no sensitivity bias is plotted as a dashed line.

H. Survey mode analysis

		Survey Mode			
	Prediction	Online	In-person	Self-report	Telephone
Vote buying	Underreporting	0	19	0	0
Turnout	Overreporting	4	3	0	3
Racial prejudice	Underreporting	5	0	0	4
Religious prejudice	Underreporting	10	0	0	2
Sexual orientation prejudice	Underreporting	11	0	4	1
	Overreporting	4	1	0	0
Support for authoritarian regimes	Underreporting	6	7	0	0
	Overreporting	1	20	0	0
	Total	41	50	4	10

Table H.8: Number of studies in each topic, by survey mode.

References

Ahart, Allison M. and Paul R. Sackett. 2004. "A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique." *Organizational Research Methods* 7(1):101–114.

Ahlquist, John S., Kenneth R. Mayer and Simon Jackman. 2014. "Alien abduction and voter impersonation in the 2012 US general election: Evidence from a survey list experiment." *Election Law Journal* 13(4):460–475.

Arentoft, Alyssa, Kathleen Van Dyk, April D. Thames, Philip Sayegh, Nicholas Thaler, Daniel Schonfeld, Joseph LaBrie and Charles H. Hinkin. 2016. "Comparing the unmatched count technique and direct self-report for sensitive health-risk behaviors in HIV+ adults." *AIDS care* 28(3):370–375.

Aronow, Peter M., Alexander Coppock, Forrest W. Crawford and Donald P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3(1):43–66.

Berinsky, Adam J. 2017. "Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *The Journal of Politics* 80(1).

Blair, Graeme and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1):47–77.

Blair, Graeme, Winston Chou and Kosuke Imai. 2019. "List Experiments with Measurement Error." *Political Analysis* 27(4):455–480.

Bøttkjær, Louise Thorn. 2017. "Crying Wolf: An Experimental Test of the Augmented List Experiment." Working paper.

Brownback, Andy and Aaron M. Novotny. 2018. "Social Desirability Bias and Polling Errors in the 2016 Presidential Election." *Journal of Behavioral and Experimental Economics* 74:38–56.

Brueckner, Hannah, Alondra Nelson and Ann Morning. 2004. "When "No" Means "Yes:" Measuring Social Desirability Effects on the Expression of Biological Concepts of Race." Time Sharing Experiments for the Social Sciences Study 275.

Cappelen, Cornelius and Tor Midtbø. 2016. "Intra-EU Labour Migration and Support for the Norwegian Welfare State." *European Sociological Review* 32(6):691–703.

Çarkoğlu, Alı and S Erdem Aytaç. 2015. "Who gets targeted for vote-buying? Evidence from an augmented list experiment in Turkey." *European Political Science Review* 7(4):547–566.

Chen, Yuyu and David Y. Yang. 2018. "The Impact of Media Censorship: Evidence from a Field Experiment in China." Working paper, Applied Economics Department, Guanghua School of Management.

Coffman, Katherine B., Lucas C. Coffman and Keith M. Marzilli Ericson. 2016. "The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated." *Management Science* 63(10):3168–3186.

Comça, Mircea and Camil Postelnicu. 2012. "Measuring social desirability effects on self-reported turnout using the item count technique." *International Journal of Public Opinion Research* 25(2):153–172.

Coppock, Alexander. 2017. "Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment." *Statistics, Politics and Policy* 8(1):29–40.

Corstange, Daniel. 2017. "Clientelism in Competitive and Uncompetitive Elections." *Comparative Political Studies* 51(1):76–104.

Coutts, Elisabeth and Ben Jann. 2011. "Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT)." *Sociological Methods & Research* 40(1):169–193.

Creighton, Mathew J. and Amaney Jamal. 2015. "Does Islam play a role in anti-immigrant sentiment? An experimental approach." *Social science research* 53:89–103.

Creighton, Mathew J., Amaney Jamal and Natalia C. Malancu. 2015. "Has opposition to immigration increased in the United States after the economic crisis? An experimental approach." *International Migration Review* 49(3):727–756.

Dai, Yaoyao. 2014. Do Citizens in Authoritarian Countries Censor Themselves? PhD thesis Duke University.

Dalton, Dan R., James C. Wimbush and Catherine M. Daily. 1994. "Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior." *Personnel Psychology* 47(4):817–829.

de Jonge, Chad P. Kiewiet and David W. Nickerson. 2014. "Artificial inflation or deflation? Assessing the item count technique in comparative surveys." *Political Behavior* 36(3):659–682.

Deming, William Edwards. 1943. *Statistical Adjustment of Data*. Wiley.

Dion, Michelle L. and Jordi Díez. 2016. "Social desirability bias and question framing effects: Evidence from a combined list and framing experiment on same-sex marriage support in Argentina." Working paper.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher and Trena M. Ezzati. 1991. "The item count technique as a method of indirect questioning: A review of its development and a case study application." *Measurement errors in surveys* pp. 185–210.

Druckman, James N., Mauro Gilli, Samara Klar and Joshua Robison. 2015. "Measuring Drug and Alcohol Use Among College Student-Athletes." *Social Science Quarterly* 96(2):369–380.

Enikolopov, Ruben, Alexey Makarin, Maria Petrova and Leonid Polishchuk. 2017. "Social Image, Networks, and Protest Participation." Working paper.

Fairbrass, Alison Julia. 2012. Using novel techniques to measure illegal killing of birds in Southern Portugal: The Unmatched Count Technique PhD thesis Department of Life Sciences, Silwood Park, Imperial College London.

Flavin, Patrick and Michael Keane. 2009. "How angry am I? Let me count the ways: Question format bias in list experiments." *Unpublished Manuscript* .

Frye, Timothy, Scott Gehlbach, Kyle L Marquardt and Ora John Reuter. 2017. "Is Putin's popularity real?" *Post-Soviet Affairs* 33(1):1–15.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Gervais, Will M. and Maxine B. Najle. 2018. "How many atheists are there?" *Social Psychological and Personality Science* 9(1):3–10.

Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77(S1):159–172.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio and David W. Nickerson. 2012. "Vote buying and social desirability bias: Experimental evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.

González-Ocantos, Ezequiel, Chad Kiewiet de Jonge and David W. Nickerson. 2015. "Legitimacy buying: The dynamics of clientelism in the face of legitimacy challenges." *Comparative Political Studies* 48(9):1127–1158.

Gosen, Stefanie. 2014. Social desirability in survey research: Can the list experiment provide the truth? PhD thesis Philipps-Universität Marburg.

Gunarathne, Anoma, Satoko Kubota, Pradeep Kumarawadu, Kamal Karunagoda and Hiroichi Kon. 2016. "Is Hiding Foot and Mouth Disease Sensitive Behavior for Farmers? A Survey Study in Sri Lanka." *Asian-Australasian journal of animal sciences* 29(2):280–287.

Haber, Noah, Guy Harling, Jessica Cohen, Tinofa Mutevedzi, Frank Tanser, Dickman Gareta, Kobus Herbst, Deenan Pillay, Till Bärnighausen and Günther Fink. 2018. "List randomization method for eliciting sexual behaviors in rural KwaZulu-Natal, South Africa." *BMC Medical Research Methodology* 18(46). Working Paper.

Heerwig, Jennifer A. and Brian J. McCabe. 2009. "Education and social desirability bias: The case of a black presidential candidate." *Social Science Quarterly* 90(3):674–686.

Hinsley, Amy, Ana Nuno, Martin Ridout, Freya A.V. St John and David L. Roberts. 2017. "Estimating the Extent of CITES Noncompliance among Traders and End-Consumers; Lessons from the Global Orchid Trade." *Conservation Letters* 10(5):602–609.

Holbrook, Allyson L. and Jon A. Krosnick. 2009. "Social desirability bias in voter turnout reports: Tests using the item count technique." *Public Opinion Quarterly* 74(1):37–67.

Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494):407–416.

Janus, Alexander L. 2005. "The List Experiment as an Unobtrusive Measure of Attitudes Toward Immigration and Same-Sex Marriages." Time Series Experiments for the Social Sciences Study 297.

Janus, Alexander L. 2010. "The Influence of Social Desirability Pressures on Expressed Immigration Attitudes." *Social Science Quarterly* 91(4):928–946.

Jayasuriya, Dinuk and John Gibson. 2013. "Elephants, tigers and safety in post-conflict Sri Lanka." CAMA Working Paper Series, Australian National University.

Kalinin, Kirill. 2014. "A Study of Social Desirability Bias in the Russian Presidential Elections, 2012." Working paper.

Kalinin, Kirill. 2015. "Exploring Putin's Post-Crimean Supermajority." Working paper.

Kalinin, Kirill. 2016. "The social desirability bias in autocrat's electoral ratings: evidence from the 2012 Russian presidential elections." *Journal of Elections, Public Opinion and Parties* 26(2):191–211.

Kalinin, Kirill and Sasha de Vogel. 2016. "Measuring Propensity of Individual Anti-Government Protest Behavior in Autocracies." Working paper.

Kiewiet de Jonge, Chad P. 2015. "Who lies about electoral gifts? Experimental evidence from Latin America." *Public Opinion Quarterly* 79(3):710–739.

Kim, Seung Hyun and Sangmook Kim. 2016. "National culture and social desirability bias in measuring public service motivation." *Administration & Society* 48(4):444–476.

Kirchner, Antje. 2013. Techniques for asking sensitive questions in labor market surveys PhD thesis Ludwig-Maximilians-Universität München.

Kramon, Eric. 2016. "Where is vote buying effective? Evidence from a list experiment in Kenya." *Electoral Studies* 44:397–408.

LaBrie, Joseph W. and Mitchell Earleywine. 2000. "Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique." *Journal of Sex Research* 37(4):321–326.

Lax, Jeffrey R., Justin Phillips and Alissa F. Stollwerk. 2016. "Are Survey Respondents Lying About their Support for Same-Sex Marriage? Lessons from A Recent List Experiment." *Public Opinion Quarterly* 80(2):510—533.

Li, Xiaojun, Weiyi Shi and Boliang Zhu. 2017. "The Face of Internet Recruitment: Evaluating the Labor Markets of Online Crowdsourcing Platforms in China." Working paper.

Lippitt, Margaret, Amelia Reese Masterson, Ana Sierra, Amy B Davis and Marney A White. 2014. "An exploration of social desirability bias in measurement of attitudes toward breastfeeding in public." *Journal of Human Lactation* 30(3):358–366.

Luks, Samantha and J. Quin Monson. 2010. "Pushing the Limits of List Experiments." Working paper, Department of Political Science, Brigham Young University.

Marques, Diogo, Ildar Muslukhov, Tiago Guerreiro, Luís Carriço and Konstantin Beznosov. 2016. Snooping on mobile phones: Prevalence and trends. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association.

Matanock, Aila M. and Miguel García-Sánchez. 2017. "Does Counterinsurgent Success Match Social Support? Evidence from a Survey Experiment in Colombia." *The Journal of Politics* 80(3).

McKenzie, David and Melissa Siegel. 2013. "Eliciting illegal migration rates through list randomization." *Migration Studies* 1(3):276–291.

Miller, Judith Droitcour. 1984. A new survey technique for studying deviant behavior PhD thesis George Washington University.

Mizuno, Elda Lorena Becerra. 2012. Does Everyone Have a Price? The Demand Side of Clientelism and Vote-Buying in an Emerging Democracy PhD thesis Duke University.

Monson, J. Quin. 2010. "Social Equality Norms for Race, Gender and Religion in the American Public During the 2008 Presidential Primaries Social Equality Norms for Race, Gender and Religion in the American Public During the 2008 Presidential Primaries." Working paper, Department of Political Science, Brigham Young University.

Morning, Ann, Hannah Brueckner and Alondra Nelson. 2019. "The expression of biological concepts of race." *Du Bois Review: Social Science Research on Race* . Forthcoming.

Prior, Markus. 2009. "Improving media effects research through better measurement of news exposure." *The Journal of Politics* 71(3):893–908.

Randrianantoandro, Tiana N., Hiroichi Kono and Satoko Kubota. 2015. "Knowledge and behavior in an animal disease outbreak–Evidence from the item count technique in a case of African swine fever in Madagascar." *Preventive veterinary medicine* 118(4):483–487.

Rayburn, Nadine Recker, Mitchell Earleywine and Gerald C. Davison. 2003. "An investigation of base rates of anti-gay hate crimes using the unmatched-count technique." *Journal of Aggression, Maltreatment & Trauma* 6(2):137–152.

Rosenfeld, Bryn, Kosuke Imai and Jacob N. Shapiro. 2016. "An empirical validation study of popular survey methodologies for sensitive questions." *American Journal of Political Science* 60(3):783–802.

Seljan, Ellen C, Todd Lochner, Maya Gold and Walker Davis. 2016. "I Know What You Did Last Cycle: Improving the Detection of State Campaign Finance Violations." *Election Law Journal* 15(3):197–213.

Sheppard, Sean C. and Mitch Earleywine. 2013. "Using the unmatched count technique to improve base rate estimates of risky driving behaviours among veterans of the wars in Iraq and Afghanistan." *Injury prevention* 19(6):382–386.

Simpser, Alberto. 2017. "Why Do Sensitive Survey Questions Elicit Truthful Answers? Theory and Evidence with Application to the RRT and the List Experiment." Working paper.

Sniderman, Paul M., Henry E. Brady and Philip E. Tetlock. 1994. "1994 Multi-Investigator Study." Dataset.

Starosta, Amy J. and Mitch Earleywine. 2014. "Assessing base rates of sexual behavior using the unmatched count technique." *Health Psychology and Behavioral Medicine: an Open Access Journal* 2(1):198–210.

Tang, Wenfang. 2016. *Populist authoritarianism: Chinese political culture and regime sustainability.* Oxford University Press.

Thomas, Alyssa S., Michael C. Gavin and Taciano L. Milfont. 2015. "Estimating noncompliance among recreational fishers: Insights into factors affecting the usefulness of the Randomized Response and Item Count Techniques." *Biological Conservation* 189:24–32.

Thomas, Kathrin, David Johann, Sylvia Kritzinger, Carolina Plescia and Eva Zeglovits. 2017. "Estimating Sensitive Behavior: The ICT and High-Incidence Electoral Behavior." *International Journal of Public Opinion Research* 29(1):157—171.

Traunmüller, Richard, Sara Kijewski and Markus Freitag. 2017. "The Silent Victims of Wartime Sexual Violence: Evidence from a List Experiment in Sri Lanka." Working paper.

Tsuchiya, Takahiro, Yoko Hirai and Shigeru Ono. 2007. "A study of the properties of the item count technique." *Public Opinion Quarterly* 71(2):253–272.

Usmani, Wali Ahmed, Diogo Marques, Ivan Beschastnikh, Konstantin Beznosov, Tiago Guerreiro and Luís Carriço. 2017. Characterizing Social Insider Attacks on Facebook. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Acm pp. 3810–3820.

Weghorst, Keith R. 2015. "Political Attitudes and Response Bias in Semi-Democratic Regimes." Working paper.

Wimbush, James C. and Dan R. Dalton. 1997. "Base rate for employee theft: Convergence of multiple methods." *Journal of Applied Psychology* 82(5):756–763.

Zigerell, L. J. 2011. "You Wouldn't Like Me When I'm Angry: List Experiment Misreporting." *Social Science Quarterly* 92(2):552–562.