The Statistical Analysis of Misreporting on Sensitive Survey Questions

Online Appendix

A. Standard list experiment model

The regression-based model for list experiments proposed by Imai (2011) models the response (Y_i^*, Z_i^*) through the use of two sub-models. The first sub-model models the probability of an affirmative response to the sensitive item, Z_i^*:

$$g(x;\delta) = \Pr(Z_i^* = 1 | X_i = x; \delta), \tag{A1}$$

where X_i is a vector of covariates and δ is a vector of parameters to be estimated. The second sub-model models the probability of an individual's response to the control items, Y_i^*:

$$h(y|x, z; \psi) = \Pr(Y_i^* = y | X_i = x, Z_i^* = z; \psi), \tag{A2}$$

where X_i is a vector of covariates, Z_i^* is the latent response to the sensitive item, and ψ is a vector of parameters to be estimated.

To derive the likelihood function, note that there are four distinct response types to the list experiment. First, the response to the control items, Y_i^*, is fully observed for respondents in the control group: those assigned to the control group are only asked to provide a response to the control items, and therefore $Y_i^* = Y_i$ if $T_i = 0$. Second and third, for respondents in the treatment group who answer $Y_i = 0$ or $Y_i = J + 1$, the response to the sensitive item Z_i^* is fully observed: responding affirmatively to none of the $J + 1$ items ($Y_i = 0$) or responding affirmatively to all of them ($Y_i = J + 1$) indicates that all individual items, including the sensitive item, were answered 0 or 1 respectively. Lastly, among those who are assigned to the treatment group and whose response is greater than 0 and less than $J + 1$, responses to both the sensitive item and control items are latent. The observed-data likelihood can be derived from recognition of these

Table A1: Respondent types for standard list experiment

Observed variables		**Latent variables**		**Observed-data likelihood**
T_i	Y_i	Y_i^*	Z_i^*	
0	Y_i	Y_i	0 or 1	$g(X_i;\delta)h(Y_i|X_i,1;\psi) + (1-g(X_i;\delta))h(Y_i|X_i,0;\psi)$
1	$J+1$	J	1	$g(X_i;\delta)h(J|X_i,1;\psi)$
1	0	0	0	$(1-g(X_i;\delta))h(0|X_i,0;\psi)$
1	$0 < Y_i < J+1$	$0 < Y_i^* < J+1$	0 or 1	$g(X_i;\delta)h(Y_i-1|X_i,1;\psi) + (1-g(X_i;\delta))h(Y_i|X_i,0;\psi)$

This table presents the individual-level likelihoods for responses to a list experiment.

response types, as shown in Table A1. The observed-data likelihood is given by the following:

$$
\begin{aligned}
L(\delta,\psi;\{T_i,Y_i,X_i\}_{i=1}^{n}) = \prod_{i=1}^{n} & \left\{g(X_i;\delta)h(Y_i|X_i,1;\psi)+(1-g(X_i;\delta))h(Y_i|X_i,0;\psi)\right\}^{1-T_i} \\
& \times\left\{g(X_i;\delta)h(Y_i-1|X_i,1;\psi)\right\}^{\mathbf{1}\{Y_i=J+1\}T_i} \times \left\{(1-g(X_i;\delta))h(Y_i|X_i,0;\psi)\right\}^{\mathbf{1}\{Y_i=0\}T_i} \\
& \times\left\{g(X_i;\delta)h(Y_i-1|X_i,1;\psi)+(1-g(X_i;\delta))h(Y_i|X_i,0;\psi)\right\}^{\mathbf{1}\{0<Y_i<J+1\}T_i}.
\end{aligned}
\tag{A3}
$$

For optimization, Imai (2011, 410-411) proposes and implements an expectation-maximization (EM) algorithm (Blair et al., 2016), details of which can be found in Imai (2011) and Blair and Imai (2012).

Similar to the goal of the main article, Blair and Imai (2012) also propose examining social desirability bias by fitting two separate list-experiment and direct-question regression models and using the fitted models to generate predicted probabilities to compare social desirability bias across groups (for details, see Blair and Imai, 2012, 54 & 60-62). This procedure is different from that outlined in the main article in that it does not seek to explicitly model inconsistency in respondents' answers.

B. Proposed method for alternative case

In the following, I lay out the proposed method for the case in which *not* answering affirmatively to the sensitive item is to provide the socially unacceptable response. For this case, we can identify three response patterns to the list experiment and direct question that define the response (Z_i^*, D_i). Substantively, these responses define the following respondent types: (1) those who hold the sensitive belief and misreport it when asked directly ($Z_i^* = 0$, $D_i = 1$); (2) those who hold the sensitive belief and do not misreport it ($Z_i^* = 0$, $D_i = 0$); and (3) those who do not hold the sensitive belief and do not misreport it ($Z_i^* = 1$, $D_i = 1$). These response patterns and their respective descriptions are presented in Table A2 (analogous to Table 1 in the main article).

The individual likelihoods for each response pattern are provided in Table A3, where the

Table A2: Respondent types defined by the response (Z_i^*, D_i) for the case in which responding $Z_i^* = 0$ is to provide the sensitive response

Type	Sensitive Z_i^*	Direct D_i	Misreport U_i^*	Description
Misreport sensitive	0	1	1	Respondent holds the sensitive belief but misreports it when asked directly.
Truthful sensitive	0	0	0	Respondent holds the sensitive belief and states so truthfully when asked directly.
Non-sensitive	1	1	0	Respondent does not hold the sensitive belief and states so truthfully when asked directly.

Note that the respondent type for the response ($Z_i^* = 1$, $D_i = 0$) is undefined by the monotonicity assumption.

observed-data model likelihood is the product of the relevant individual likelihoods as given in the last column of the table. Models $g(\cdot)$, $j(\cdot)$, and $h(\cdot)$ remain as defined in the main body of the article. In this alternative case, note that by the monotonicity assumption, $j(x,1,t;\gamma)=0$.

For the EM algorithm, the complete-data likelihood function is given by the following:

$$\begin{aligned} L(\delta,\gamma,\psi;\{T_i,Y_i,X_i,Z_i^*,U_i^*\}_{i=1}^n) = \prod_{i=1}^n & \Big\{g(X_i;\delta)\{1-j(X_i,1,T_i;\gamma)\}h(Y_i-T_i|X_i,1,0;\psi)\Big\}^{\mathbf{1}(Z_i^*=1\wedge U_i^*=0)} \\ & \times\Big\{\{1-g(X_i;\delta)\}\{1-j(X_i,0,T_i;\gamma)\}h(Y_i|X_i,0,0;\psi)\Big\}^{\mathbf{1}(Z_i^*=0\wedge U_i^*=0)} \\ & \times\Big\{\{g(X_i;\delta)\}j(X_i,0,T_i;\gamma)h(Y_i|X_i,0,1;\psi)\Big\}^{\mathbf{1}(Z_i^*=0\wedge U_i^*=1)}. \end{aligned} \tag{A4}$$

The expressions used to calculate the weights for the EM algorithm are given in Table A4.

The maximization step computes the parameters δ, γ, and ψ using the most recent values of the weights, $w_i^{(\cdot)}$, from the E-step, that maximize the observed-data log-likelihood function given as follows:

$$\begin{aligned} & Q(\delta,\gamma,\psi;\{Y_i,X_i,T_i,Z_i^*,U_i^*,w_i^{(\cdot)}\}_{i=1}^n) = \\ & \quad \sum_{i=1}^n w_i^{(\text{non-sensitive})}\Big\{\log g(X_i;\delta)+\log\{1-j(X_i,1,T_i;\gamma)\}+\log h(Y_i-T_i|X_i,1,0;\psi)\Big\} \\ & + \ w_i^{(\text{truthful sensitive})}\Big\{\log\{1-g(X_i;\delta)\}+\log\{1-j(X_i,0,T_i;\gamma)\}+\log h(Y_i|X_i,0,0;\psi)\Big\} \\ & \quad + \ w_i^{(\text{misreport sensitive})}\Big\{\log\{1-g(X_i;\delta)\}+\log j(X_i,0,T_i;\gamma)+\log h(Y_i|X_i,0,1;\psi)\Big\}. \end{aligned} \tag{A5}$$

C. Screener question

The screener question was asked as follows:

```
The previous question contained a list of statements.
Which of the following subjects was a part of that list?
```

- `The power of unions`
- `Gay marriage`
- `The federal budget`
- `Don't know`

Respondents who did not answer "The power of unions" (7% of respondents) were removed from the dataset used in the results section.

I also check for the possibility that treatment assignment affects responses to the screener question. It may be the case that receiving a slightly longer list (5 instead of 4 items) could lead respondents to be less attentive because of the added effort and length of time required to complete the question. In both the treatment and control group however, 7% of respondents did not respond to the screener question correctly. The difference between the two groups (0.5 percentage points) is not statistically significant (p = 0.12).

Table A3: Observed-data likelihood for the case in which $Z^* = 0$ is the socially undesirable response

Observed variables			Latent variables			Observed-data likelihood
T_i	Y_i	D_i	Z_i^*	U_i^*	Y_i^*	
0	Y_i	1	0 or 1	0 or 1	Y_i	$g(X_i;\delta)(1-j(X_i,1,0;\gamma))h(Y_i\|X_i,1,0;\psi)+(1-g(X_i,\delta))j(X_i,0,0;\gamma)h(Y_i\|X_i,0,1;\psi)$
0	Y_i	0	0	0	Y_i	$(1-g(X_i;\delta))(1-j(X_i,0,0;\gamma))h(Y_i\|X_i,0,0;\psi)$
1	$0<Y_i<J+1$	1	0 or 1	0 or 1	Y_i-1 or Y_i	$g(X_i;\delta)(1-j(X_i,1,1;\gamma))h(Y_i-1\|X_i,1,0;\psi)+(1-g(X_i,\delta))j(X_i,0,1;\gamma)h(Y_i\|X_i,0,1;\psi)$
1	0	0	0	0	0	$(1-g(X_i,\delta))j(X_i,0,1;\gamma)h(0\|X_i,0,1;\psi)$
1	$J+1$	0	—	—	—	Violates monotonicity assumption
1	$Y_i<J+1$	0	0	0	Y_i	$(1-g(X_i;\delta))(1-j(X_i,1,1;\gamma))h(Y_i\|X_i,0,0;\psi)$
1	0	1	0	1	0	$(1-g(X_i;\delta))j(X_i,0,1;\gamma)h(0\|X_i,0,1;\psi)$

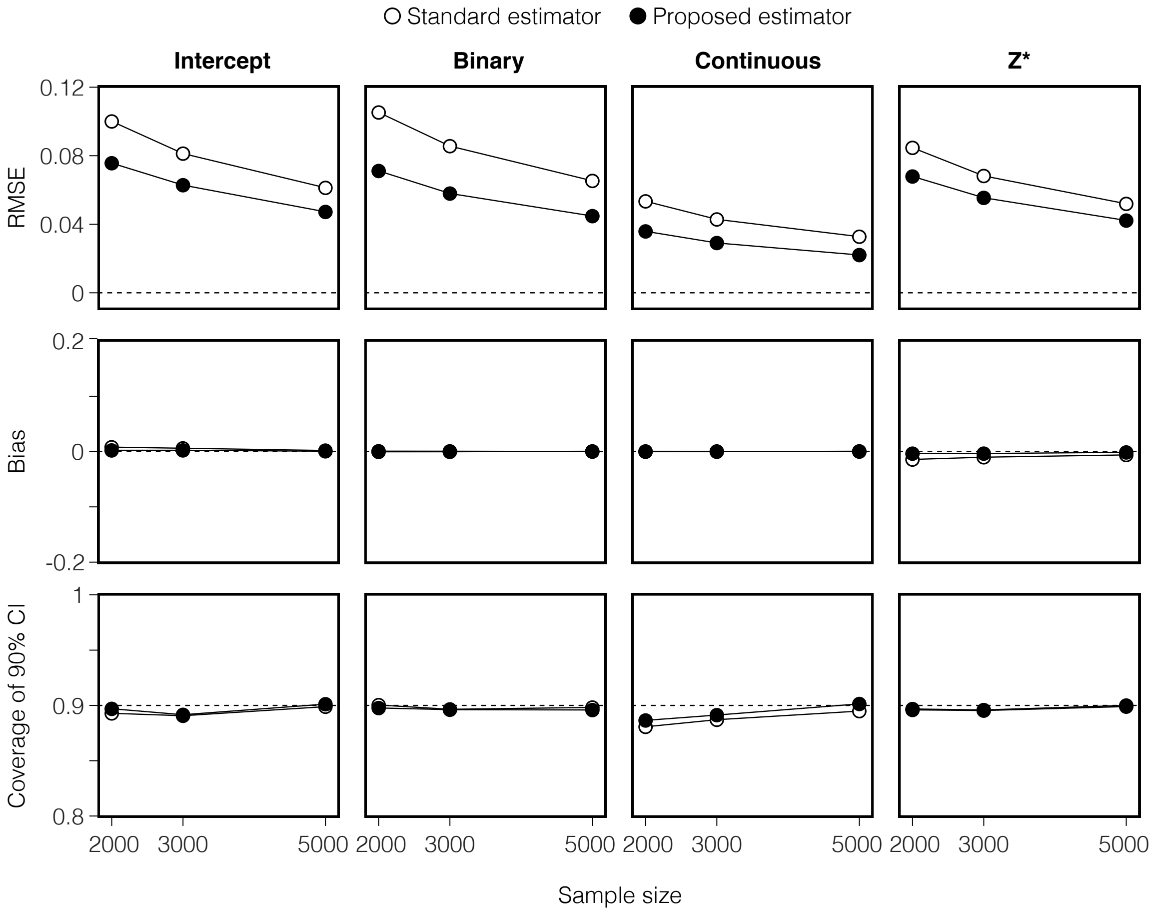
Table A4: Weights $w_i^{(\cdot)}$ as calculated in the E-step for the case in which $Z^* = 0$ is the socially undesirable response

$$w_i^{(\text{non-sensitive})} = \frac{g(X_i;\delta)\{1-j(X_i,1,T_i;\gamma)\}h(Y_i-T_i|X_i,1,0;\psi)}{g(X_i;\delta)\{1-j(X_i,1,T_i;\gamma)\}h(Y_i-T_i|X_i,1,0;\psi)+\{1-g(X_i;\delta)\}\{1-j(X_i,0,T_i;\gamma)\}h(Y_i|X_i,0,0;\psi)+\{1-g(X_i;\delta)\}j(X_i,0,T_i;\gamma)h(Y_i|X_i,0,1;\psi)}$$

$$w_i^{(\text{truthful sensitive})} = \frac{\{1-g(X_i;\delta)\}\{1-j(X_i,0,T_i;\gamma)\}h(Y_i|X_i,0,0;\psi)}{g(X_i;\delta)\{1-j(X_i,1,T_i;\gamma)\}h(Y_i-T_i|X_i,1,0;\psi)+\{1-g(X_i;\delta)\}\{1-j(X_i,0,T_i;\gamma)\}h(Y_i|X_i,0,0;\psi)+\{1-g(X_i;\delta)\}j(X_i,0,T_i;\gamma)h(Y_i|X_i,0,1;\psi)}$$

$$w_i^{(\text{misreport sensitive})} = \frac{\{1-g(X_i;\delta)\}j(X_i,0,T_i;\gamma)h(Y_i|X_i,0,1;\psi)}{g(X_i;\delta)\{1-j(X_i,1,T_i;\gamma)\}h(Y_i-T_i|X_i,1,0;\psi)+\{1-g(X_i;\delta)\}\{1-j(X_i,0,T_i;\gamma)\}h(Y_i|X_i,0,0;\psi)+\{1-g(X_i;\delta)\}j(X_i,0,T_i;\gamma)h(Y_i|X_i,0,1;\psi)}$$

Figure A1: Control-items sub-model for Simulation Study 2

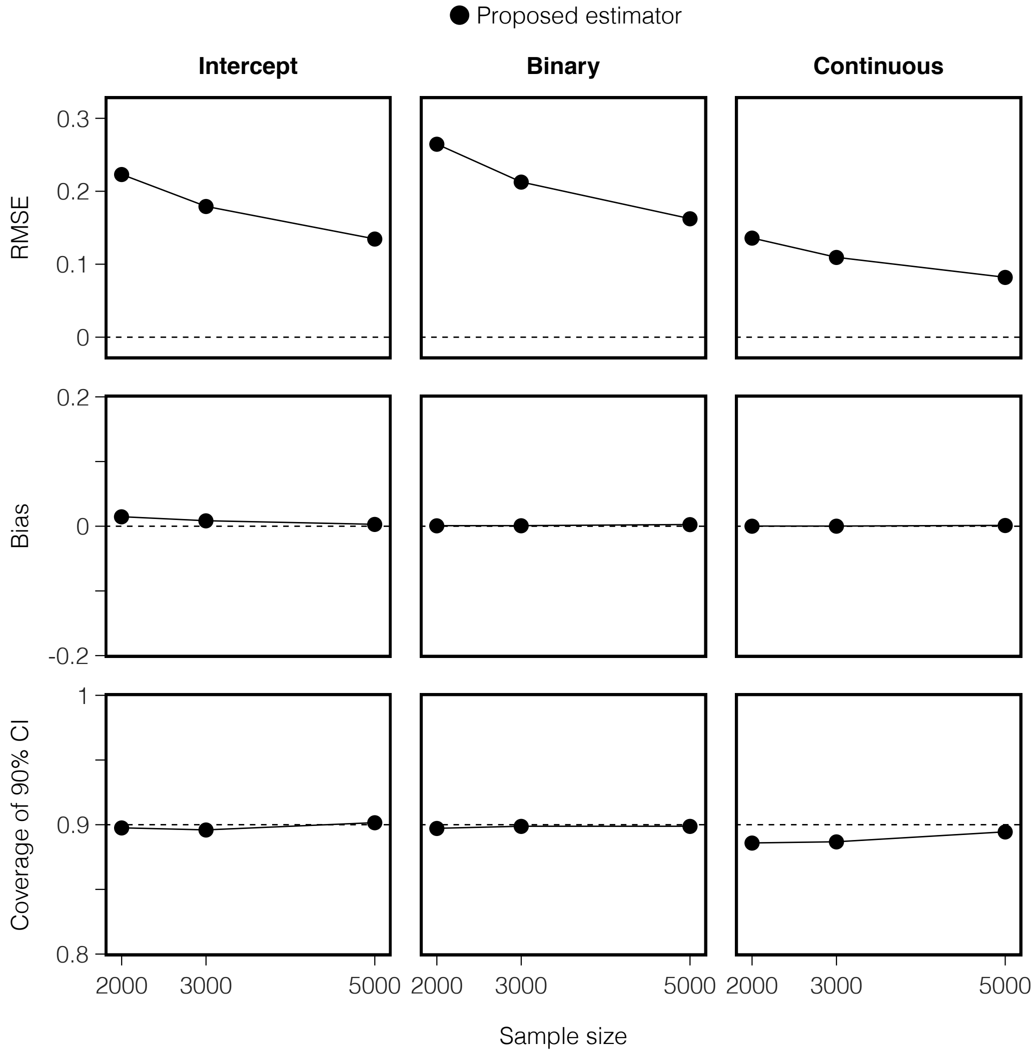


This figure presents data from 10,000 Monte Carlo simulations for root mean squared error (RMSE), bias, and confidence interval coverage for the control-items sub-model from Simulation Study 2.

D. Control-items and misreport sub-model results from Simulation Study 2

Figures A1 and A2 present results from Simulation Study 2 (section 3.2 of the main article). These figures compare the results from the proposed to the standard estimator is terms of RMSE, bias, and coverage of confidence intervals. Although the control-items sub-model is rarely, if ever, of substantive interest to researchers, we can see that the proposed estimator improves on the standard estimator in terms of RMSE. Note that Figure A2 does not include data from any simulations for the standard estimator because it does not include a misreport sub-model.

Figure A2: Misreport sub-model for Simulation Study 2



This figure presents data from 10,000 Monte Carlo simulations for root mean squared error (RMSE), bias, and confidence interval coverage for the misreport sub-model from Simulation Study 2.

E. List experiment control-items sub-model for empirical application

Table A3 presents results from the control-items sub-models for the models presented in Table 4 in the main article. The parameters $Z^* = 1$ and $Z^* = 0$ represent indicator variables for whether a respondent answered affirmatively (or not) to the sensitive item. U^* denotes an indicator variable for whether someone who holds the sensitive belief (i.e. $Z^* = 0$) also misreports it.[1]

[1] Note that *not* responding affirmatively to the statement "Women are as capable as men in politics" is the sensitive response.

Table A3: Multivariate list experiment results (control-items sub-model)

	Model 1 Control items		**Model 2 Control items**	
	Coef	SE	Coef	SE
Ideology				
Ideology (0 = right, 10 = left)	0.023***	(0.003)	0.019***	(0.003)
Gender				
Male (*baseline*)				
Female	-0.035*	(0.015)	-0.035*	(0.015)
Age group				
Age 18-29 (*baseline*)				
Age 30-39	-0.059*	(0.024)	-0.059*	(0.024)
Age 40-49	-0.074**	(0.026)	-0.078**	(0.026)
Age 50-64	-0.168***	(0.022)	-0.176***	(0.022)
Age 65+	-0.111***	(0.023)	-0.111***	(0.023)
Education				
High school or below (*baseline*)				
College	-0.036	(0.025)	-0.031	(0.025)
University degree	0.168***	(0.022)	0.169***	(0.022)
Mother tongue				
English (*baseline*)				
French	0.023	(0.027)	0.024	(0.026)
Other language	0.012	(0.025)	0.015	(0.024)
Region				
Ontario (*baseline*)				
Atlantic	-0.094**	(0.035)	-0.104**	(0.035)
Quebec	-0.028	(0.025)	-0.032	(0.025)
West	-0.052**	(0.019)	-0.052**	(0.019)
$U^* = 1$	—	—	0.406***	(0.076)
$Z^* = 0$	—	—	-0.093	(0.048)
$Z^* = 1$	-0.283***	(0.058)	—	—
Constant	-0.258***	(0.059)	-0.509***	(0.034)
N	22,372		22,372	

This table presents results from the list experiment control-items sub-models for the standard and proposed estimator for the sensitive statement "Women are as competent as men in politics." $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

F. Tests for design-assumption violations in empirical application data

A series of preliminary checks are run on the data used in the empirical section to test for violations of the list experiment's design assumptions. I first test for a "design effect," which refers to a difference in responses to the control items (Y_i^*) associated with treatment assignment (for details, see Blair and Imai, 2012, 63-65). The test for the presence of a design effect proposed by Blair and Imai (2012) shows no strong evidence of one (p = 1).

Second, I check for violations of the monotonicity assumption. Violations of this assumption are immediately apparent for respondents in the treatment group who respond that they agree with none of the items in the list experiment question, but answer affirmatively to the direct question. Such a response pattern would indicate that the socially unacceptable response is given in the list experiment, but the socially acceptable response is given to the direct question. Three out of 11,133 respondents in the treatment group provide this response pattern and are removed from the dataset. I then compare list experiment responses among those who openly admit to holding the socially unacceptable response. Among the group of respondents who provide this response ("No") to the direct question, the difference in the mean response to the list experiment question between treatment and control groups should be 0 in expectation: these respondents will, by the monotonicity assumption, not respond affirmatively to the sensitive item in the list experiment and therefore responses to the list experiment should not depend on treatment assignment.[2] Testing for a difference in the mean response to the list experiment question between the control and treatment group for those who provide the socially unacceptable response to the direct question does not provide strong evidence of an assumption violation ($p = 0.07$).

Lastly, I test whether those who are assigned to the treatment group respond systematically differently to the direct question than those in the control group. As noted previously, to avoid this problem, the list experiment and direct questions were separated from each other by a large number of unrelated questions. There is no strong evidence that treatment assignment affects responses to the direct question ($p = 0.77$).

References

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.

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

Blair, Graeme, Kosuke Imai, Bethany Park, Alexander Coppock and Winston Chou. 2016. "list: Statistical Methods for the Item Count Technique and List Experiment." Available at The Comprehensive R Archive Network (CRAN), https://CRAN.R-project.org/package=list.

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

[2] This test is analogous to that proposed by Aronow et al. (2015, 50-51) (Placebo Test I), which tests for a difference of 1 between control and treatment groups among those who respond affirmatively to the direct question for the case in which responding affirmatively to the sensitive item is to provide the socially unacceptable response.