The Size of the LGBT Population and the Magnitude of Anti-Gay Sentiment are Substantially Underestimated[†]

Katherine B Coffman
Ohio State University

Lucas C Coffman
Ohio State University

Keith M Marzilli Ericson
Boston University SOM
and NBER

January 26, 2016

Abstract

We demonstrate widely-used measures of anti-gay sentiment and the size of the LGBT population are misestimated, likely substantially. In a series of online experiments using a large and diverse but non-representative sample, we compare estimates from the standard methodology of asking sensitive questions to measures from a "veiled" methodology that precludes inference about an individual but provides population estimates. The veiled method increased self-reports of *anti*-gay sentiment, particularly in the workplace: respondents were 67 percent more likely to disapprove of an openly gay manager when asked with a veil, and 71 percent more likely to say it should be legal to discriminate in hiring on the basis of sexual orientation. The veiled methodology also produces larger estimates of the fraction of the population that identifies as LGBT or has had a sexual experience with a member of the same sex. Self-reports of non-heterosexual identity rose by 65 percent, and same-sex sexual experiences by 59 percent. We conduct a "placebo test" and show that for non-sensitive placebo items, the veiled methodology produces effects that are small in magnitude and not significantly different from zero in seven out of eight items. Taken together the results suggest anti-gay discrimination might be a more significant issue than formerly considered, as the non-heterosexual population and anti-gay workplace-related sentiment are both larger than previously measured.

[†] This paper previously circulated under the title "Privacy is not Enough: The Size of the LGBT Population and the Magnitude of Anti-Gay Sentiment are Substantially Underestimated".

1.   Introduction

When analyzing lesbian, gay, bisexual, or transgender (LGBT)-related sentiment, managers and policy-makers typically have to rely on self-reported answers to questions such as, 'do you consider yourself homosexual', or 'are you comfortable with an openly gay manager.' Such data can affect managerial decisions[1] and public policy. Yet answers to these questions might be biased towards social norms; respondents might prefer to give socially approved answers rather than honest answers. Thus, widely-used data from surveys and polls may not be accurate, and changes in measured LGBT-related sentiment could in part be due to changes in reporting. Our understanding of the degree of animus towards LGBT groups might be mis-calibrated, affecting both managerial practice and social policy.

In this paper, we report the results of an experiment that tests whether anonymity and privacy are sufficient for eliciting truthful responses to questions about sexuality— i.e., whether current best practices eliminate social desirability bias. We find substantial underreporting of LGBT identity and behaviors as well as underreporting of anti-gay sentiment even under anonymous and very private conditions.

Knowing the prevailing norms, the extent of anti-LGBT sentiment, and how many are likely affected, is an important first step in creating healthy work environments. Current data on LGBT sentiment seem to be suggesting great improvements. In a 2013 Pew nationally representative survey of the (self-reported) LGBT community, 92% agreed society is now more accepting compared to ten years ago (Pew 2013). However, in the same survey, over one fifth claimed they had been discriminated against at work. Moreover, the policy issue most commonly cited as a top priority for LGBT respondents was equal employment rights, rather than marriage or adoption. Even in the absence of outright discrimination, a rift between beliefs or expectations of how employees feel about LGBT issues and how they honestly feel could create tension.

Having accurate measures of sexual orientation and behaviors is also important for research and policy. Many areas of research use data about the LGBT population. For instance, it is used to study discrimination in the labor market[2] and the value of urban amenities,[3] as well as sexually-transmitted diseases and policies to reduce them.[4] Data on LGBT individuals has been used to test theories of the economics of the family, including household labor supply, educational investment, the demand for

---

[1] See Li and Nagar (2013) on the LGBT diversity in the workplace, and Klawitter and Flatt (1998) on the effect of antidiscrimination policies.
[2] For a review of the economics of LGBT families, see Black, Sanders, and Taylor (2007). Black, Makar, Sanders, and Taylor (2003) examine the gay male wage "penalty" and the lesbian "premium." See also Ahmed and Hammarstedt (2010), Allegretto and Arthur (2001), Clain, and Leppel (2001), Jepsen (2007), Weichselbaumer (2003). See Badgett (2001) for a review.
[3] Black, Gates, Sanders, and Taylor (2002) examine the location decisions of the LGBT population.
[4] See Bloom and Glied (1992), Berg and Lien (2006), Black, Gates, Sanders and Taylor (2000), and Fay, Turner, Klassen and Gagnon (1989).

children, and the gender-based divisions of labor.[5] Data on the LGBT population is used by many academic fields, as well as by firms and social service organizations. Similarly, data on LGBT-related sentiment affects not only the policy choices of the government, but also those of firms and non-profit organizations.

The results in this paper add to a burgeoning literature measuring discrimination. The economics literature on discrimination has typically avoided asking about beliefs directly, instead relying on the observation of behavior, and has largely focused on race and gender discrimination. For instance, discrimination has been identified using field experiments that have featured auditors attempting to purchase cars (e.g. Ayres & Siegelman 1995), fake resumes submitted to potential employers (Bertrand & Mullainathan 2004), emails asking for an academic meeting (Milkman, Akinola, and Chugh 2012), and strategic games with subtle racial identification (Fershtman and Gneezy 2001). Our approach is to document the existence of bias in beliefs. Previous evidence suggests that discriminatory beliefs are linked discriminatory behavior.[6] In recent work, Glover, Pallais, and Pariente (2015) link managers' unconscious bias to employee performance, illuminating a direct link between levels of bias and firm outcomes. Moreover, direct evidence on beliefs can illuminate how preferences map into behavior, and augment existing evidence in many ways—for instance, by examining perceptions of social norms, and by distinguishing between taste-based versus statistical discrimination.

Individuals are reluctant to respond honestly on surveys in a variety of contexts because they may prefer their answer to adhere to social norms—a phenomenon known as "social desirability bias" (Maccoby and Maccoby 1954, Edwards 1957, Fisher 1993).[7] Moreover, in certain contexts, individuals may fear direct harm from disclosing certain information if it is not kept confidential. As a result, behaviors, beliefs, or identities that could be perceived as sensitive or unpopular are typically underreported. Social norms regarding LGBT-related issues have changed rapidly in recent years. As a result, we do not know the extent to which underreporting is a problem for LGBT-related topics; in some cases, it is not obvious which direction individuals would distort their answers.

Surveys about LGBT-related issues have been improving in a variety of ways. Survey researchers have shown that truthful reporting increases with *anonymity* (not being able to link an individual's responses to her identity) and *privacy* (not being able to observe an individual while she gives her responses) (Das and Laumann 2010, Office of National Statistics 2008, Ellison and Gunstone

---

[5] See Carpenter (2007) on educational investment in college, Jepsen and Jepsen (2002) on assortative mating, Oreffice (2011) on household bargaining, and Lundberg and Pollak (2007) for a review.
[6] See Greenwald et al. (2009) for a review of how unconscious and conscious beliefs link to behavior, and Dasgupta and Rivera (2006) for evidence regarding sexual orientation in particular.
[7] See Kuran (1995) for a related analysis of "preference falsification".

1998). As a result of these advances, recent data on sexual orientation from well-worded and well-executed surveys have been reported with some confidence (Chandra, Mosher, Copen and Sionean 2011). However, it is unknown how accurate these data are.

We utilize a method designed to reduce social desirability bias, the item count technique (ICT) (Miller 1984).[8] (It is also known as the "unmatched count" or "list response" technique.) The ICT is a between-subject method in which a randomly chosen control group of participants is asked to report how many of N items are true for themselves.[9] The rest of the respondents report how many of N+1 items are true, with N of those items being identical to the control group's items, and the N+1st item being the sensitive item of interest, e.g. "I am not heterosexual". With a large enough sample, the researcher can estimate the population mean for the N+1st item of interest by differencing out the mean of the sum of the N other items as estimated from the control group.

Using this design, a researcher can never perfectly infer an individual's answer to the sensitive item, so long as a respondent does not report that either 0 or N+1 items are true. The veil provided by the ICT thus all but eliminates precise inference about an individuals' answer to the item of interest (though the researcher may still make probabilistic statements). We make a modification to the traditional ICT that not only allows for correct inference at the population level, but also allows us to estimate the survey population's rate of misrepresentation under traditional survey methods. Our control group sees the list of N statements and reports how many are true. Immediately following, they are asked the sensitive item directly. We refer to this condition as the "Direct Report" treatment. The second group, the "Veiled Report" treatment, sees the N+1 items as in the traditional ICT.

Using this modification, we test whether questions relating to sexual orientation are stigmatized— do they show evidence of social desirability bias even when asked in a self-administered, computer-assisted survey? We find evidence that many questions relating to sexual identity or related views have a substantial social desirability bias even under extreme privacy and anonymity. The veiled method increased self-reports of non-heterosexual identity by 65 percent (p<0.05), same-sex sexual experiences by 59 percent (p<0.01), and same-sex attraction by 9.4 percent (n.s.). We combine all own-

---

[8] Evidence from voting has also been used to test for the existence of social desirability bias. For instance, Powell (2013) compares the discrepancy between polling and voting for ballot initiatives, and finds that there is a larger discrepancy for same-sex marriage. This is consistent with respondents not truthfully answering polling questions about same-sex marriage. It could also be consistent with systematic misprediction in who votes elections that deal with same-sex marriage (e.g. groups that feel strongly about same-sex marriage might be better than average at organizing get-out-the-vote efforts, perhaps due to role religious communities play.) In related work, Stephens-Davidowitz (2013) finds that Obama underperformed, relative to expectation, in areas that had unexpectedly more racist Google searches.

[9] These N items can either be neutral items or sensitive items. Individuals are never asked about the N items directly in either condition.

sexuality questions into an index, and find that the Veiled Report treatment significantly raises the number of sensitive answers overall (p<0.01).

The veiled method also increased the measured rates of *anti*-gay sentiment. [10] Respondents were 67 percent more likely to express discomfort with an openly gay manager at work (p<0.01), 71 percent more likely to say it should be legal to discriminate in hiring on the basis of sexual orientation (p<0.01), 22 percent less likely to support the legality of same-sex marriage, 46 percent less likely to support adoption by same-sex couples (p<0.10), and 32 percent *less* likely to state they believe homosexuality is a choice (p<0.05). We again combine all the opinion questions into an index, and find that the Veiled Report treatment significantly raises the overall number of intolerant answers (p<0.01). Based on these results, we designed a second experiment to provide more detail on workplace-related LGBT sentiment. We find that the veiled method increases the fraction of respondents who reported being unhappy with having a gay co-worker. However, for less personal questions (legality of refusing to serve LGBT customers, appropriateness about being openly gay at work), we find smaller, statistically insignificant effects.

We also consider how norms vary across demographic subgroups. The costs of reporting an identity or belief may vary by demographic characteristics because group identities vary (Akerlof and Kranton 2000). Our point estimates suggest that individuals in demographic categories that other research has identified as more openly anti-gay—Christians, African-Americans, and older populations (Herek and Glunt 1993) –are more likely to lie about their sexual identity without a veil.

Finally, we directly investigate the validity of the ICT method more generally using a placebo test: we conduct an ICT experiment for 8 "placebo" statements that should not be affected by social desirability bias. Seven of the eight placebo items produce treatment effects that are statistically indistinguishable from zero.

We also experimentally show that the ICT works better when the sensitive answer is "no", as compared to "yes". This can be helpful for future managers designing questionnaires in determining the framing of the questions. Further, it lends more credence to the interpretation of our other results, as our weak results typically occur only when the sensitive answer was a "yes". When we used the ICT most efficiently, with a "no" as the sensitive answer, the results tell a robust story.

Our finding that anti-gay sentiment is socially undesirable is consistent with recent findings reported in the popular press and opinion polls that suggest a social norm of acceptance of the LGBT community and support for pro-LGBT policies (CNN ORC Poll 2012). And yet, in spite of that norm, we

---

[10] There may also be implicit discrimination (Bertrand & Mullainathan 2005), which the ICT does not necessarily capture.

find evidence that many individuals remain uncomfortable reporting non-heterosexual identity and behavior.

## 2.	Existing Literature

2.1 Existing Measures of the LGBT Population

Research on the LGB[11] population in the U.S. has been hindered due to lack of data availability, as few representative surveys ask about sexual orientation. Surveys vary in many factors that can affect the fraction of the population measured as LGB ("incidence rates"): sample selection (e.g. all adults v. adults 18-44), the way questions are worded, and the degree of privacy and anonymity afforded to participants. The modern literature based on representative samples in both the U.S. and other Western countries is discussed by Gates (2011); we draw on his review below.[12] For self-identification as LGB, estimates range from 1.7 percent of adults (National Epidemiological Survey on Alcohol and Related Conditions, 2004-2005) to 5.7 percent of adults (National Survey of Sexual Health and Behavior 2009).

Other ways of measuring sexual orientation produce much higher rates. The National Survey of Family Growth (NSFG), conducted by the CDC's National Center for Health Statistics, interviewed a representative sample of adults aged 18-44 using audio computer-assisted self-interviewing, in which answers are entered into a computer rather than spoken to the researcher. (See Chandra et al. 2011 for detail on this survey.) In it, 11 percent of adults reported any same-sex attraction, and 8.8 percent of adults reported any same-sex sexual behavior; the fraction of adults identifying as LGB was 3.7 percent in that survey.

More recently, the Pew Research Center (2013) attempted to survey a representative sample of people who identify as LGBT. The results illustrate the difficulty of identifying the LGBT population—only about half of LGBT people who responded to this survey say that all or most of the important people in their life are aware they are LGBT. They also illustrate the importance of further research on the LGBT population's economic and other life outcomes: as a result of their sexual orientation or gender identity, 21 percent say they have been discriminated against by an employer and 30 percent say they have been physically attacked or threatened.

2.2 Validation of the ICT

---

[11] The existing literature has treated the lesbians, gays, and bisexuals separately from transgender individuals. However, our paper typically refers to the LGBT population, since many transgender individuals may not identify as heterosexual in our questions. In the best estimates of Gates (2011), 3.5 percent of the population identifies as LGB and only 0.3 percent as transgender.

[12] The early work of Kinsey, Pomeroy and Martin (1948) was not based on a representative sample of the population.

In a variety of contexts, the ICT has been shown to elicit more reports of behaviors that may be perceived as socially undesirable (Tourangeau and Yan 2007, Blair and Imai 2012). It generally increases respondents' perception of privacy, as compared to other computer-aided elicitations (Coutts and Jann 2011). Blair, Imai and Lyall (2014) provide validation of the ICT by showing it gives similar results to a very different method for reducing social desirability bias ("endorsement experiments").

The ICT has been used to examine a variety of behaviors, including voter turnout (Holbrook and Krosnick 2010), employee theft (Dalton, Wimbush, and Daily 1994), and the incidence of sexuality-related hate crimes on a college campus (Rayburn, Earleywine, and Davison 2003). It has also been used to study patterns of sexual behavior, including risky sexual behaviors and alcohol abuse (LaBrie and Earleywine 2000), sexual experiences with same-sex partners among high school students in Miami (Zimmerman and Langer 1995), and risky sexual practices among Ugandans (Jamison, Karlan, and Raffler 2013). The ICT has never been applied to measure sexual orientation of the general population, and rarely to measure opinions about public policy.

Previous research has documented that the ICT provides increased estimates of prevalence *only* for stigmatized behaviors. Put differently, it is not the case that increased reporting under the veil of the ICT is simply mechanical. Tsuchiya et al. (2007) reports the results of a placebo test of the ICT; while they find that the ICT produces an increase in 10 percentage points in reporting of a stigmatized behavior (shoplifting), they find no significant increase in reporting of an innocuous behavior (blood donation). We provide similar evidence in our own placebo tests, described in Section 6.

The ICT method is related to other ways of preventing individual level inference for sensitive survey questions. Most notable is the randomized response technique (RRT), in which respondents use a private randomization device (i.e. flip a coin) to determine whether they answer either a sensitive or innocuous question. The RRT has been shown to successfully elicit more sensitive answers across contexts than direct questioning (Lensvelt-Mulders et al 2005). However, the RRT can be more difficult to implement online, and subjects trust the RRT less than the ICT (Coutts & Jann 2011). In addition, recent research by John et al. (2013) has demonstrated that participants may not respond to the randomization device relied upon by the RRT as instructed, in an attempt to avoid appearing as though they provided the sensitive response. With the ICT, the answer to the sensitive question is completely veiled for the vast majority of participants (those who do not respond that 0 or N+1 items are true), minimizing the incentive to misrepresent.

3.   Experiment Design

In our main experiment, we investigate eight questions, detailed in Table 1. Three questions deal with participants' sexuality: whether they consider themselves heterosexual, whether they are sexually

attracted to members of the same sex, and whether they have had a sexual experience with someone of the same sex.[13] The remaining five questions examine attitudes and opinions related to sexuality—participants are asked about public policy issues, such as legal recognition of same-sex marriage, as well as personal beliefs and feelings, such as being comfortable with LGBT individuals in the workplace. For reporting convenience only, we will define the potentially "sensitive answer" as the answer that would disclose non-heterosexuality (for own sexuality questions) or anti-gay opinions (for opinion questions). This definition does not affect how we conducted the analysis (we use all two-sided tests), nor was it presented to participants.

Participants took the survey online on their own computers (giving them privacy from the researcher) and never disclosed identifying information (anonymity). Participants first answered demographic questions. Then, to assure understanding of the elicitation, we provided all participants with an example of how to respond to a list using only non-sensitive items.

Participants' answers to eight potentially sensitive, sexuality-related questions were elicited under two randomly assigned treatments, "Direct Report" or "Veiled Report" (treatments are assigned across-subject, and each participant answers all eight questions within a given treatment). The Direct Report treatment was designed not only to serve as a control treatment, but also to replicate common existing survey designs, in which participants must respond directly to a sensitive question. The Veiled Report treatment was based on the ICT methodology and allowed the participant to provide truthful information about the sensitive question without disclosing it to the researcher. To enable the Veiled Report treatment, each sensitive question was paired with four other items. We used a different set of four items for each question, but the sets used and which questions they were paired with were held constant across treatment.

For two reasons, each set of four items were composed of two pairs of items we selected to be negatively correlated. First, the negative correlation reduces variance in the sum of the sensitive items, increasing our statistical power (see Glynn 2013 for a discussion on this point). Second, the negative correlations also decrease the likelihood that either zero or five items are true for a respondent in the Veiled Report treatment, ensuring that we cannot make inferences about sensitive topics at the individual level.

---

[13] While all of our questions have binary (yes/no) answers, this is not a claim that sexual orientation, attraction, or even discomfort are binary concepts. They are indeed appropriately measured on a spectrum. Our questions merely ask whether the respondent considers themselves to be at a particular point on the spectrum. Our hypothesis of interest is about across-treatment differences, rather than levels, and this restriction to a binary space seems unlikely to impact the size of our estimated treatment effects.

In the Direct Report condition, participants first saw a list of four statements and were asked to indicate how many of the four statements were true for them.[14] Then, they were asked to respond directly to the sensitive question, "Yes" or "No". In the Veiled Report treatment, participants saw a list containing the four statements and the sensitive item, rephrased in statement format. They were then asked to indicate how many of the five statements were true for them. Subjects were randomly assigned to treatments after completion of the demographic questions. This allowed us to stratify according to age. Participants were classified as belonging to one of three age brackets: 30 years of age and under, 31-50, 51 and over. Within each bracket, participants were randomly assigned to the Direct Report or the Veiled Report in equal proportions. In addition, participants were randomly assigned to one of two order conditions, either answering the sensitive questions in the order listed in Table 1, Panel B or in the reverse ordering.[15] The order of statements within each question was the same for all subjects.

Following the questions of interest, all participants answered a question on risk preferences, completed the cognitive reflection task (CRT), and were asked to submit their zip code. We used this as a check of attention. Since subjects provided their state of residence in the demographic section, we can match up the zip code and state to check for consistency. Table A2 in the Appendix shows that our estimated treatment effects are not significantly changed by restricting our analysis to the participants who we measure to be consistent on this dimension.

4.   Empirical Approach

For each question q and participant i in the Veiled Report treatment, we observe y_{qi}^V, the number of the five statements reported as true. In the Direct Report treatment, we observe d_{qi}, equal to one if participant i answered "yes" to the directly asked sensitive question and zero otherwise and c_{qi}, the number of the four statements reported as true. For the Direct Report treatment, we construct the sum of these measures, $y_{qi}^D = d_{qi} + c_{qi}$, which gives the number of five items reported as true for the participant in the Direct Report treatment.

---

[14] The directly asked item appears after the four-item list in the Direct Report treatment. In the Appendix, we report the results from a follow-up experiment that documents that answers to a directly asked sensitive question do not vary with whether or not it is presented with the four-item list..

[15] In Table A1, we investigate the impact of order assignment on participant responses to Questions 1 and 8. Consistent with previous literature, we find that participants who saw Question 1 last, rather than first, were marginally more likely to reveal the sensitive answer, non-heterosexual identity. This seems to be true across both the Direct and Veiled Report treatments. We also see some evidence that answering the sensitive questions in reverse order—that is, answering Question 8 first— reduced attrition in both treatments (see Appendix), though attrition overall was less than 3 percent. This may be because participants perceived Question 8 as less sensitive than Question 1.

Table 1. Experimental Design

Panel A: Comparison of Direct Report and Veiled Report Treatments

Direct Report	**Veiled Report**
• I remember where I was the day of the Challenger space shuttle disaster.	• I remember where I was the day of the Challenger space shuttle disaster.
• I spent a lot of time playing video games as a kid.	• I spent a lot of time playing video games as a kid.
• I would vote to legalize marijuana if there was a ballot question in my state.	• I would vote to legalize marijuana if there was a ballot question in my state.
• I have voted for a political candidate who is pro-life.	• I have voted for a political candidate who is pro-life.
	• I consider myself to be heterosexual.

Please fill in the bubble that corresponds to the total number of statements above that apply to you.

Direct Report: 0   1   2   3   4

Veiled Report: 0   1   2   3   4   5

Do you consider yourself to be heterosexual?

Yes   No

Panel B: Sensitive Questions Used

	Question	Sensitive Answer
Own Sexuality		
1. Heterosexual	Do you consider yourself to be heterosexual?	"No"
2. Attraction	Are you sexually attracted to members of the same sex?	"Yes"
3. Experience	Have you had a sexual experience with someone of the same sex?	Yes"
LGBT-related Sentiment		
4. Marriage	Do you think marriages between gay and lesbian couples should be recognized by the law as valid, with the same rights as heterosexual marriages?	"No"
5. Manager	Would you be happy to have an openly lesbian, gay, or bisexual manager at work?	"No"
6. Discriminate	Do you believe it should be illegal to discriminate in hiring based on someone's sexual orientation?	"No"
7. Adopt	Do you believe lesbians and gay men should be allowed to adopt children?	"No"
8. Change	Do you think someone who is homosexual can change their sexual orientation if they choose to do so?	"Yes"

Under truthful reporting, the expected number of true items should be the same in the two conditions since participants are randomly assigned: $E[y_{qi}^V] = E[y_{qi}^D]$. However, when they differ, $E[y_{qi}^V]$ is a better estimate of the true population mean under the assumption that the Veiled Report treatment lowers the cost of telling the truth.

We define the change in reporting[16] as $\mu_q \equiv E[y_{qi}^V] - E[y_{qi}^D]$. We can also interpret μ as a measure of how stigmatized the sensitive response is; a larger μ suggests the existence of a social norm which makes truthful reporting of the sensitive answer in the Direct Report treatment more costly.

Rather than simply comparing sample means, regression analysis gives a better and more precise estimate of μ_q, as it allows us to control for observed demographics. Thus, in our results below, we will report the estimated μ from the regression:

$$y_{qi} = \beta X_i + \mu_q V_i$$

where $V_i \in \{0,1\}$ is an indicator variable for being in the Veiled Report treatment, and y_{qi} is simply y_{qi}^V or y_{qi}^D, whichever is observed for the individual.

The vector of observed demographic controls X_i includes age (linearly and as a quadratic), education (some high school, high school graduate, some college, college graduate, some graduate school, finished graduate school), political affiliation (Republican, Democrat, independent/other), religion (Christian, Jewish, no religion, other), race (white, black, other), gender (male, female, transgender), census region (Midwest, West, South, Northeast), marital status (single, married, other), religiosity (on a scale of 1-7), and political engagement (on a scale of 1-7).

5.   Experiment Results

We recruited participants from an online labor market, Amazon's Mechanical Turk (MTurk). Previous studies have shown that this subject population is culturally and demographically diverse (Paolacci, Chandler, and Iperiotis 2010) and displays similar behavior in experiments to standard samples (Rand 2011; Horton, Rand, and Zeckhauser 2010). Data for the main experiment were collected in two waves. The first wave was conducted from November 1st – November 3rd, 2012, just prior to the United States' presidential election.[17] 786 individuals participated over this three-day window. The second wave

---

[16] In our design, our hypothesized sensitive answer to the question varies—sometimes it is "Yes" and other times "No". In our analysis, we recode the data so that a positive change in reporting indicates an increase in the hypothesized sensitive answer, and a negative change indicates a decrease.

[17] The United States presidential election motivated our use of the two-wave design. Same-sex marriage appeared as a ballot question in four states: Maine, Maryland, Minnesota, and Washington. In addition to identifying the main treatment effect of the Veiled Report condition, we also were interested in exploring post-election differences in reported opinions on LGBT issues, particularly in these battleground states. We did not have data to do power calculations for these proposals ex ante; after Wave 1,

was conducted just after the presidential election, from November 7th – November 15th, 2012; 1730 individuals participated during this window.[18]

Our sample is diverse, with a broad range of demographic characteristics, but it is not a representative sample: it is younger, more educated, and more liberal than the U.S. general population. Table A5 in the appendix provides descriptive statistics. Our sample is approximately 42 percent female with a median age of 26. Less than 32 percent describe themselves as being at least moderately religious, and less than 16 percent self-reports as Republican. Attrition in the experiment was very low (2.97 percent of participants assigned to treatment), and did not differ significantly by treatment.[19] The median time spent by participants was 5.27 minutes.

Because the sample is non-representative, the focus will be exclusively on across-treatment differences and percentage changes in reporting, rather than on the levels of behaviors or opinions. Generally, however, the groups we under-sample are groups we estimate to have relatively larger treatment effects. Hence, if the treatment effects differed between our sample and a representative sample, our data suggest the representative sample might show an even larger effect of reporting method. Below, we present results for the full sample. If we analyze only the subsample for which we infer high levels of attention and/or thoughtfulness in their responses, our results are not qualitatively changed (see Appendix).

Before turning to our regression results, we first present the histograms of responses to each of our questions. In Figure 1, we graph the distributions of y_{qi} for each question. There are a few important observations to draw from the histograms. First, fewer than 7 percent of participants are at the boundaries (0 or 5) for any particular question. This assures that our choice of items did in fact provide an effective veil for a large majority of our sample; we cannot infer the truth about the sensitive item at the individual level for 93 percent of the sample. Second, the distributions of y_{qi} are very clearly non-uniform. More

---

we determined that we would only have power to identify the main treatment effect, not changes in the treatment effect in these states post-election.

[18] The design of the Direct Report treatment was slightly different in the first wave. Within each question, one other statement was separated from the list and asked directly. This difference is illustrated in Table A3. Our intention was to obfuscate the purpose of the study, drawing attention away from the fact that each directly asked question was sexuality-related. However, if individuals respond differently to the other item when it is asked directly, this may confound our estimation efforts for the sensitive item. Therefore, the design was altered for the second wave. This change does not affect our results: in Table A4, we show that our estimated treatment effects are similar for both waves of elicitations, though there is less precision in each smaller subsample. The only difference in estimated treatment effects occurs for Question 2 Same-Sex Attraction, where treatment effects are large and significant in the first wave but not the second.

[19] 2667 individuals began the survey. 74 of them did not complete the first demographics screen (which was common across treatments.) Of the 2593 individuals who saw the first treatment screen, 2516 (97%) completed the entire experiment. Of those in the Direct Report treatment, 45 attrited, while 32 attrited from the Veiled Report treatment. Attrition is thus 1 percentage point higher in the Direct Report treatment. Under the most conservative assumption that all of these additional attriters would have given the sensitive answer had they stayed in the experiment, the treatment effects in column 2 of Table 3 would be reduced by only 1 percentage point.

importantly, if we compare the distributions across treatments for any particular question, they look much more similar than if we compare the distributions across question. Taken together, these observations suggest that our participants are responding in an informative manner to our elicitation. We address the issue of inattentive response in more detailed in Section 6.

Table 2 presents our primary results. Column 1 shows the percent reporting the sensitive answer in the Direct Report treatment. Column 2 shows the change in reporting, μ, as a percent of the total sample, estimated using a regression with controls described in Section 4.[20] (Note that μ that has been recoded so that it gives the increase in reporting of the sensitive answer.) Column 3 estimates the percent, in this sample, for whom the sensitive answer is true, and is derived by adding Columns 1 and 2. Column 4 gives the percent increase in reporting of the sensitive answer under the Veiled Report; it is derived by dividing Column 2 by Column 1.

We present heteroskedasticity-robust standard errors for the treatment effect μ in Table 2. When calculating the percent increase in respondents answering yes to the sensitive question (Table 2, column 4) and the estimated true fraction answering yes to the sensitive question (Table 2, column 3), we use the bootstrap to calculate standard errors (estimated from 1000 repetitions, stratified on treatment). The choice of method for deriving standard errors does not matter much. Bootstrap standard errors that do not stratify on treatment are very similar to the ones reported, and bootstrap standard errors for the treatment effect μ are quite similar to the heteroskedasticity-robust ones reported.

5.1 Own Sexuality Questions

For participants' own-sexuality questions, the Veiled Report treatment has a sizable impact on two of three questions. "Question 1-Heterosexual" asks whether the participant identifies as heterosexual (yes/no). We do not describe the alternative categories for non-heterosexuality, but these could encompass homosexual, gay, lesbian, bisexual, queer, undecided, and other categories.[21] In the Direct Report treatment, 11 percent of the population reports that they do not consider themselves heterosexual (8 percent for men, 16 percent for women). In the Veiled Report treatment, this increases to 19 percent (15 percent for men, 22 percent for women). The 7.3 percentage point difference is significant at p<0.05, and represents a 65 percent increase in the fraction of the sample reporting as non-heterosexual.

---

[20] A simple comparison of means, presented in Table A6, gives similar results.
[21] Our wording of the question is motivated by the existence of several plausible alternative categories. We can distinguish between non-heterosexual and heterosexual, but not among the categories encompassed by non-heterosexuality. This seemed preferable to, for instance, identifying gay or not gay, which may have led to a failure to separate between someone who considered themselves bisexual (not gay) or heterosexual (not gay). Miller (2001) finds that most survey respondents understand the term "heterosexual".

In "Question 3-Experience", the number of participants reporting having had a sexual experience with someone of the same sex increases from 17 percent (12 percent for men, 24 percent for women) in the Direct Report treatment to 27 percent (17 percent for men, 43 percent for women) in the Veiled Report treatment, a 59 percent increase (difference, $p<0.01$).

For "Question 2-Attraction", we estimate little underreporting of same-sex attraction (1 percentage point), a difference that is not statistically significantly different from zero. However, our confidence intervals cannot reject a substantial 8 percentage point increase. (In general, we may not observe a treatment effect if the cost of truth-telling is low in both conditions – that is, there is no social stigma associated with the sensitive answer --or if the cost of truth-telling is not lowered enough with the veil. Low base rates also make it difficult to identify a treatment effect, and if participants interpreted this question as being *exclusively* attracted to members of the same sex, it would drive the base rate down.[22]

Finally, we create an "own sexuality index" for each individual by summing the answers to each of the separate own sexuality questions. We do this to increase our power to detect an overall effect of our treatment. We code the questions so that positive answers indicate sensitive answers, as described in Table 1. Thus, higher values of this index indicate a greater degree of LGBT identity, experience, and/or attraction. In the "sum" version of the index, we simply sum the number of items a participant said yes to for each question. In the normalized version, we place lower weight on questions with more variance by dividing this number of yes items for each question by that question's standard deviation. The two indices are quite similar.

Table 3 reports the results using this index. Using the sum version of the index, we find an increase in the own sexuality index of 0.19 for Veiled Report condition, indicating that the total number of sensitive answers for these 3 questions is 0.19 higher with the Veiled Report than the Direct Report. A non-parametric two-sample Wilcoxon rank-sum test indicates the difference between the two conditions is significant with $p<0.01$. Using the normalized index, the estimated increase in total number of sensitive items is 0.20 ($p<0.01$).

---

[22]Participants may have interpreted this question as indicating being exclusively or primarily attracted to members of the same sex, which would have reduced levels across both treatments. (Note that Gates (2011) finds that a majority of individuals who identified as LGBT considered themselves bisexual.) In a separate survey, also conducted on Mechanical Turk, we asked 72 individuals from a population similar to our sample to predict how likely various types of individuals would be to answer "Yes" to this question. The results indicate that bisexual or bi-curious individuals would be less likely to answer "Yes" to this question, which would not be expected if participants interpreted the question as asking whether they are "at all attracted" to members of the same-sex. The results of that survey can be found in Figure A1 in the Appendix.

Heterosexual

Attraction

Experience

Marriage
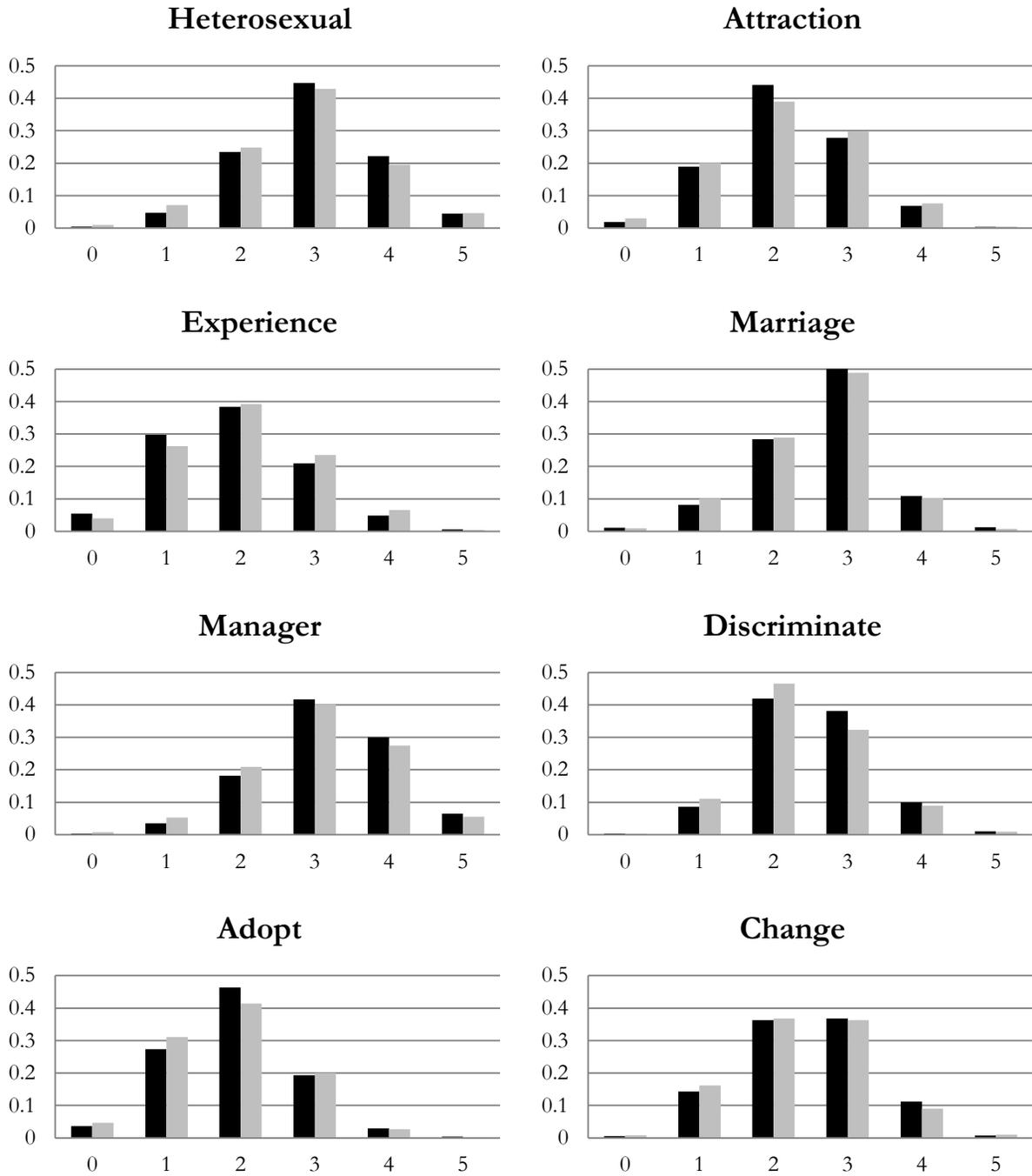
Manager

Discriminate

Adopt

Change

Figure 1. Distributions of Total Number of Yeses to Each Question

Note: Veiled Report in light gray, Direct Report in black. The x-axis gives the number of yeses reported, and the y-axis gives the fraction of sample that reported that number of yeses.

Table 2: The Effect of Veiled Report Treatment on Reports of Sensitive Behaviors

Sensitive Answer	Percent Reporting Sensitive Answer, Direct Report	ΔReporting of Sensitive Answer, Veiled Report	Estimated True Fraction for Sensitive Answer	Percentage Increase in Sensitive Answer, Veiled Report
Own Sexuality				
Not Heterosexual	11.3	7.3	18.6	64.2
	[0.89]	[3.57]	[3.54]	[33.2]
Same sex Attraction	13.9	1.3	15.3	9.5
	[0.97]	[3.63]	[3.57]	[26.9]
Same-sex sexual experience	17.2	10.1	27.4	58.7
	[1.06]	[3.82]	[3.75]	[24.1]
LGBT-related sentiment				
Not Support Same-sex Marriage	18.8	4.2	23.0	22.5
	[1.10]	[3.18]	[3.08]	[17.4]
Not Happy with LGB Manager	16.2	10.8	27.0	66.6
	[1.03]	[3.75]	[3.72]	[24.7]
Not Illegal to Discriminate	14.4	10.3	24.7	71.7
	[0.99]	[3.33]	[3.39]	[25.7]
LGB not allowed to Adopt	12.9	5.9	18.8	45.9
	[0.94]	[3.40]	[3.36]	[27.4]
Can Change Orientation	22.2	-7.0	15.2	-31.4
	[1.17]	[3.52]	[3.39]	[15.5]

Notes: n= 2516, with 1270 in Direct Report condition. Each row corresponds to a separate regression. Column 1 is the sample mean of respondents in the Direct Report condition. Column 2 is the coefficient μ on "Veiled Report" from a regression with controls. All answers coded such that positive numbers reflect increases in reporting of sensitive responses. Column 3 adds column 1 and 2, while Column 4 divides Column 2 by Column 1. Standard errors in brackets: Column 2 presents heteroskedasticity-robust standard errors, Columns 3 and 4 standard errors are derived using the bootstrap.

Table 3: Veiled Report Treatment Effect: Indices

	Number of Sensitive Answers per Subject			
	LGBT Identity Index		Anti-Gay Sentiment Index	
	Sum	Normalized	Sum	Normalized
Treatment Effect	0.187	0.196	0.243	0.279
	[0.0657]	[0.0700]	[0.0827]	[0.0934]
R^2	0.132	0.131	0.183	0.181

Notes: *n= 2516, with 1270 in Direct Report condition. Normalized index sums the number answered to each question, divided by the standard deviation of that question in the Direct Report treatment. Treatment Effect is the coefficient μ on "Veiled Report" from a regression with controls. Heteroskedasticity-robust standard errors in brackets.

5.2 LGBT-related Sentiment

Next, we examine attitudes and opinions related to sexual orientation. The evidence suggests participants underreport anti-LGBT sentiment when asked directly. In "Question 4-Marriage", 19 percent of the Direct Report treatment did *not* support the legal recognition of same-sex marriages. This increases to 23 percent in the Veiled Report treatment. (This 4 percentage point difference is not statistically significant from zero.)

The veiled treatment has the largest impact on reported attitudes toward LGBT individuals in the workforce. In "Question 5-Manager", the percent of the population that would *not* be happy to have a LGBT manager at work increases by 69 percent in the Veiled Report treatment compared to the Direct Report treatment, from 16 percent to 27 percent (p<0.01). "Question 6-Discriminate", asks whether the respondent believes it should be illegal to discriminate in hiring based upon sexual orientation. While only 14 percent in the Direct Report treatment say that this type of discrimination should *not* be illegal, in the Veiled Report treatment, we estimate that 25 percent of our sample believes it should not be illegal (difference, p<0.01).

Adoption by LGBT couples has received less media attention than same-sex marriage, but is still the subject of an ongoing debate, with state laws varying in the degree to which they permit LGBT couples to adopt. In both conditions, a minority of our sample opposes LGBT adoption. However, opposition is stronger in the Veiled Report treatment (19 percent opposed) than in the Direct Report treatment (13 percent opposed, p<0.10).

"Question 8-Change" is somewhat different than the other sentiment questions, as it asks about a factual belief rather than an opinion on a LGBT-relevant policy. Here, participants were asked whether

they believe a person can change their sexual orientation if they choose to do so. The Veiled Report treatment *decreases* the percent reporting that sexual-orientation is changeable, from 22 percent under Direct Report to 15 percent (p<0.05). This indicates that participants saw it more socially desirable to report that sexual orientation is changeable. Ex ante, we anticipated that a response of "yes", a person cannot change their sexual orientation if they choose to do so, was the answer most closely aligned with an anti-gay norm. (Pro-LGBT groups often argue that sexual orientation is genetic, not a choice.) However, it is possible that our participants instead perceived "yes" as a pro-LGBT answer: they may have interpreted the ability to change sexual orientation as an affirmation of being free to shift along a spectrum, or being free to choose a partner of either gender.

Just as for the "own sexuality" questions, we sum the answers of the 5 sentiment questions to create an overall sentiment index. Here, the questions are coded so that positive answers indicate our hypothesized sensitive answers (anti-LGBT sentiment). Table 3 shows that, using this index, the number of sensitive anti-gay sentiment answers[23] rises by 0.24 in Veiled Report condition (two-sample Wilcoxon rank-sum test, p<0.01).

5.3 Treatment Response by Demographics

Table 4 examines the effect of the Veiled Report method on own sexuality questions, broken out by the following subgroups: gender, race, religious affiliation, political affiliation, and age. For reference, we provide the Direct Report responses in Table A7 in the Appendix. We hypothesize that our treatment effects (that is, underreporting of non-heterosexuality in the Direct Report treatment) should be larger for demographic groups with social norms that are perceived as less LGBT-friendly: Christians, older respondents, and Black/African Americans (Herek and Glunt 1993).

The data support our hypotheses. Among Christians in our sample, the Veiled Report condition raises reports of non-heterosexuality by 13 percentage points (from 8 percent to 21 percent) in Question 1 (p<0.05) and same-sex sexual experiences by 14 percentage points (from 11 percent to 25 percent) in Question 3 (p<0.05), compared to the Direct Report. These are increases of 163 percent and 127 percent, respectively. Among participants with no religious affiliation, the Veiled Report treatment produces much smaller differences in these questions (point estimates: 0.02 and 3.7 percentage points, respectively).

The effect of the Veiled Report method is also larger for older individuals. For a subsample of participants 31-50 years of age, the percent identifying as non-heterosexual increases from 9 percent to 30 percent (p<0.01), a 233 percent increase, and the fraction reporting a same-sex sexual experience

---

[23] Note that because Question 8 goes in the opposite direction for the other sentiment questions, its treatment effect actually reduces the treatment effect measured for the index as a whole.

increases from 18 percent to 38 percent (p<0.05), a 111 percent increase. In contrast, the Veiled Report treatment has no impact on reporting about own sexuality among individuals who are 30 and younger in our sample. Though our sample size is too small to reliably estimate racial differences, our point estimates indicate that the Veiled Report treatment had a larger effect among Blacks/African Americans than Whites in our sample.

In Table A9, we present the treatment effects for Questions 4-8 by demographic groups (Averages for Direct Report responses can be found in Table A8.) There are fewer striking differences in treatment effects across demographic groups for opinions on LGBT issues. The model predicts that the Veiled Report should have a stronger impact on those for whom the costs of deviating from the social norm are largest. The social norm of support for LGBT rights is likely stronger among Democrats than Republicans, and so conforming with it may be more socially desirable for Democrats.

Question 4-Marriage deals with perhaps the most politically-polarized LGBT policy issue. The estimated fraction of Republicans who do *not* support the legal recognition of same-sex marriage increases by 6 percentage points (48 percent in the Direct Report, 54 percent using the Veiled Report), an insignificant difference. For Democrats, the treatment effect is larger, with our estimate of non-supporting Democrats increasing from 10 percent to 20 percent using the Veiled Report (p<0.05).

Turning to Questions 5-Manager, 6-Discriminate, and 7-Adoption, results vary by religious affiliation, with stronger treatment effects for Christians than those with no religious affiliation. The Veiled Report treatment has a significant impact on both Democrats and Republicans for the employment questions (5 and 6), but the magnitude of the effects are larger for Republicans than Democrats. The estimated fraction of Republicans who report that they would not be happy with an LGB manager at work nearly doubles, going from 35 percent to 67 percent (p<0.01). When asked directly, only 23 percent of Republicans in our sample report that it should *not* be illegal to discriminate in hiring based upon sexual orientation; our estimated fraction with Veiled Report more than doubles to 47 percent (p<0.01). These results suggest that, unlike in the case of same-sex marriage, the belief that it is socially unacceptable to be intolerant of LGBT individuals in the workplace may be widely-shared by nearly all demographic groups.

Table 4: Change in Sensitive Answer Reports for Own Sexuality Questions in Veiled Report, by Demographics

	1-NonHetero.	2-Attract.	3-Experience	N		1-NonHetero.	2-Attract.	3-Experience	N
Gender					*Politics*				
Male	6.59	4.55	4.69	1444	Democrat	13.3	0.618	5.54	1155
	[4.46]	[4.82]	[4.91]			[5.33]	[5.25]	[5.64]	
Female	6.30	2.32	18.9	1058	Republican	3.34	12.2	0.889	400
	[5.9]	[5.63]	[6.13]			[9.39]	[9.36]	[9.36]	
Race					*Age*				
White	6.98	2.13	9.48	2022	Under 31	3.3	0.476	4.43	1658
	[3.94]	[4.06]	[4.28]			[4.09]	[4.52]	[4.66]	
Black	22.9	3.15	23.6	151	31 – 50	20.9	4.53	19.6	700
	[17.6]	[14.9]	[18.1]			[7.67]	[6.84]	[7.52]	
Religion					51 plus	9.22	3.79	32.3	158
Christian	12.9	4.51	13.8	905		[15.7]	[18.7]	[17.1]	
	[6.46]	[6.03]	[6.25]						
No Religion	0.0204	5.01	3.71	1078					
	[4.87]	[5.55]	[5.85]						

Note: n= 2516. The table gives the coefficient μ on "Veiled Report" from a regression with controls run on each demographic subgroup, equivalent to Column 2 of Table 2. (Details in Appendix). Heteroskedasticity-robust standard errors in brackets.

5.4 Additional Workplace-Related LGBT Sentiment

Understanding LGBT-related sentiment has management implications—bias is likely to influence workplace culture, employee and customer interactions, and employee satisfaction, potentially impacting firm productivity. Table 2 showed the most hidden anti-LGBT related sentiment for the two questions most of interest to managers – would you be happy with an openly gay boss, and should it be legal to discriminate on the basis of sexual orientation. Thus, we designed an additional wave of the experiment to obtain more detail on LGBT-related sentiment, particularly in the workplace.

The structure of the experiment was similar to that in the first wave. Conducted in April 2015, N= 2881 participants first faced the same demographic questions as in wave 1, and then were randomly assigned to Private or Direct Ask treatments. They then answered some placebo questions (discussed in Section 6). Then, they saw four different questions about LGBT-related sentiment, with results shown in Table 5.

The results indicate that reported opinions about LGBT-related sentiment should be interpreted cautiously. There seemed to be little distortion in opinions about the appropriateness of refusing to serve LGBT customers and the appropriateness of being openly gay at work. However, we found substantial distortion about a more personal question: participants were unwilling to directly admit to being unhappy about working with a gay coworker (similar to our previous result about being unhappy working with a

gay manager). Finally, we find some suggestion of hidden pro-LGBT sentiment in a question involving children: participants are more likely to say that LGBT individuals should avoid working closely with children when asked directly than when asked privately.

In Q9, we asked "Do you think it should be legal for small businesses to refuse to serve gay and lesbian clients?" Whether and how LGBT individuals are protected by non-discrimination clauses have been the subject of public debate and lawsuits. In our Direct Report condition, only 19.8% of participants said refusing service should be legal, and there was no statistically significant change in the Veiled Report (the point estimate is a 4.4 percentage point reduction in the fraction saying discrimination should be legal).

In Q10, we asked "Do you think it is wrong for gays and lesbians to be open about their sexual orientation at work?" Understanding employee opinions is relevant for managers crafting diversity policy and complying with antidiscrimination law. (In July 2015 the Equal Employment Opportunity Commission ruled that the Civil Rights Act of 1964 prohibited discrimination based on sexual orientation, though there is continued litigation in the courts.) In our Direct Report condition, only 9.4% of participants said that openness about sexual orientation was wrong, and there was no statistically significant change in the Veiled Report condition, with a point estimate of nearly 0.

In Q11, we asked "Would you be happy to work closely with an openly gay or lesbian co-worker?" This question is very similar to Q5 from the first wave, which about an openly gay or lesbian manager. Both the levels of agreement for both questions are similar, as well as the changes from the Veiled Report. In the Direct Report condition, about 84% of participants in the first wave said they would be happy with a gay manager, and about 91.3% of participants here said they would be happy with a gay coworker. The Veiled Report condition reduced the fraction who said they would be happy with a gay coworker by 9.7 percentage points (p<0.01), very similar to the 10 percent reduction in participants who would be happy with a gay manager in the first wave.

Finally, in Q12, we asked "Would you prefer openly gay, lesbian, or transgender individuals avoid working closely with children?" LGBT teachers working in schools have faced discrimination and dismissal, and until July 2015, Boy Scouts banned LGBT adults from volunteering with the organization. We find that 18% of the sample, when asked directly, says they would prefer LGBT individuals from working closely with children. However, in the Veiled Report condition, we see a significant 9.32 percentage point reduction in the fraction of the sample wanting LGBT individuals to avoid working with children. (p<0.01). The direction of these results perhaps suggests that individuals wanted to appear "protective" of children, but were in fact comfortable with LGBT individuals working with children.

Table 5. Workplace Related LGBT-Related Sentiment

Hypothesized Sensitive Answer	Percent Reporting Sensitive Answer, Direct Report	ΔReporting of Sensitive Answer, Veiled Report	Estimated True Fraction for Sensitive Answer	Percentage Increase in Sensitive Answer, Veiled Report
Should be Legal to Refuse Service to LGBT	19.8	-4.4	15.5	-22.0
	[1.05]	[3.36]	[3.24]	[16.6]
Wrong to be Open about Orientation at Work	9.4	0.3	9.1	2.99
	[0.77]	[2.96]	[2.88]	[3.23]
Unhappy with Gay Coworker	9.7	9.7	19.4	99.3
	[0.78]	[3.12]	[3.07]	[34.7]
Prefer LGT Individuals to Not Work with Children	18.0	-9.32	8.7	-51.7
	[1.02]	[3.21]	[3.01]	[16.9]

Notes: *n= 2881 with 1438 in the Direct Report condition. Column 1 is the sample mean of respondents in the Direct Report condition. Column 2 is the coefficient μ on "Veiled Report" from a regression with controls. Column 3 adds column 1 and 2, while Column 4 divides Column 2 by Column 1. Standard errors in brackets: Column 2 presents heteroskedasticity-robust standard errors, Columns 3 and 4 standard errors are derived using the bootstrap.

6.   The Validity of the ICT Mechanism: Robustness Checks

In this section, we explore alternative stories for the across treatment differences we document. We tackle this issue in three ways. First, we run a placebo test of our method and show that we do not estimate significant treatment effects for the Veiled Report method for seven out of eight non-sensitive items. Second, we explore models of inattentive response and document that our data are inconsistent with the patterns these models would predict. Third, we explore whether the effectiveness of the ICT depends upon the framing of the sensitive answer.

6.1 Placebo Test

We document significant increases in the rate of reporting of sensitive behaviors and opinions under our Veiled Report treatment as compared to the Direct Report treatment. We attribute these increases to the lower cost of truth-telling in the Veiled Report treatment. In this section, we explore the validity of this interpretation by investigating whether our Veiled Report treatment has a significant impact on reporting for non-sensitive, placebo items. If we observe that the Veiled Report treatment *also* produces a difference in estimated prevalence of non-sensitive behaviors and opinions, we might worry

that our effects are produced not by an increase in truthful reporting but rather by something mechanical. Here, we show that this does not appear to be the case.

We test eight placebo items in two additional studies. Each of these items can be found in Table 8. Our goal was to construct prompts that would not generate any social desirability bias. We aimed to maintain as much similarity with our sensitive items as possible; thus, each placebo item is a question about the participant that the researcher does not know the answer to. The only difference is that we believe none of the possible answers to these questions should carry any stigma.

Participants were recruited through Amazon Mechanical Turk in two waves, with four placebo items tested in each wave. These waves were conducted in April 2015 and July 2015. Each participant began the experiment by completing the same demographic survey that participants in the original study completed. Then, participants faced four list questions with placebo statements.[24] We used the same list items that we employed in our first study. This allows to compare, for a given set of list items, the treatment effect produced using a placebo item of interest to the treatment effect using a sensitive item of interest.[25]

In Table 6, we display the effect of the Veiled Report treatment on the percent of participants who report "Yes" to each placebo item. Note that unlike our sensitive items of interest, there is no natural "sensitive" answer for these placebo items; e.g. we do not have an ex ante hypothesis about whether short or long sleeves are more socially desirable. Thus, we simply report changes in the percentage of "yes" responses for each question. Column 1 contains the percentage of "yes" responses in the Direct Report treatment and Column 2 contains the estimated percentage point increase in the percent of participants responding "yes" to the placebo question under the Veiled Report condition. Finally, in Column 3, we display the estimated percentage point increase in the percent of participants responding "yes" to the sensitive item under the Veiled Report condition when we used the *same* list items but paired with a sensitive item of interest (our results from Study 1). Thus, if we expect that it is our list items or some other aspect of the Veiled and Direct Report treatments that produce the treatment effects, we would expect to see very similar treatment effects (both direction and magnitudes) in Columns 2 and 3. On the other hand, if we see different treatment effects depending upon whether the item of interest is sensitive

---

[24] They then saw four additional statements: in April 2015, we examined questions about workplace-related LGBT sentiment, as previously discussed, and in July 2015 we examined question phrasing, as discussed in Section 6.3.

[25] We re-use six of our eight original lists. We use the six that produced the largest treatment effects in the expected direction in our original study (Questions 1, 3, 4, 5, 6, and 7) in an attempt to understand whether these large treatment effects could be driven by something mechanical. We use the Question 5 list twice, as the first placebo item we tested using it produced a significant treatment effect. We also made up one new list to pair with placebo item 8.

or not, we can feel more confident that the treatment effects we observe for our LGBT questions are indeed due to the fact that the Veiled Report treatment reduces the cost of truth-telling.

For the seven placebo items that were paired with lists we had previously used with sensitive items, six produce treatment effects that are directionally smaller in magnitude than the original treatment effect using the same list with a sensitive item. For seven of our eight total placebo items, we find no significant effect of the Veiled Report on the rate of reporting of "yes" to the item of interest. For only one placebo item, Question 3, "Does your birthday fall on an odd-numbered day?", we find a significant decrease in the rate of reporting of "yes" under our Veiled Report treatment (11.3 percentage points, $p<0.01$). Because we found a significant effect for this placebo item using this list (the list for Question 5, "Manager" from the original study), we re-used this list in our second wave of placebo testing, pairing it with a new placebo item, Question 6, "Have you worn contact lenses before?". In this second placebo test, we found no significant treatment effect (a decrease of 1.19 percentage points, $p=0.65$).

Taken together, we do not have strong evidence that the effects found when using the ICT with sensitive items are due to a systematic bias of the mechanism. Using the same lists, with the same population, we were not able to consistently produce significant or substantial results in a placebo condition.

Table 6. The Effect of Veiled Report Treatment on Reporting of Placebo Items

Placebo Item	Percent Reporting "Yes", Direct Report	Δ Percent Reporting of *"Yes" on Placebo Item*, Veiled Report	Δ Percent Reporting of *"Yes" on Sensitive Item* using Same List, Veiled Report	Question Number of Original Sensitive Item that Used Same List
1. Are you completing this survey from a laptop computer?	61.2	-3.52	-7.3**	1 – Heterosexual
	[1.29]	[3.54]	[3.57]	
2. Are you wearing a long-sleeved shirt right now?	26.8	4.15	10.1***	3 – Experience
	[1.17]	[3.81]	[3.82]	
3. Does your birthday fall on an odd-numbered day?	52.6	-11.3***	-10.8***	5 – Manager
	[1.32]	[3.65]	[3.75]	
4. Are you wearing a wristwatch right now?	10.2	1.67	-10.3***	6 - Discriminate
	[0.77]	[3.22]	[3.33]	
5. Are you a Verizon customer?	32.8	-3.06	-4.2	4 – Marriage
	[0.87]	[2.37]	[3.18]	
6. Have you worn contact lenses before?	45.3	-1.19	-10.8***	5 – Manager
	[0.92]	[2.65]	[3.75]	
7. Has it rained once where you live in the last four days?	71.7	-0.94	-5.9*	7 – Adopt
	[0.83]	[2.38]	[3.40]	
8. Are you wearing a blue shirt right now?	18.7	-4.54	N/A	N/A
	[0.72]	[2.79]		

Notes: *n= 2881 for Qs 1 – 4, n=5828 for Qs 5 – 8. Column 1 is the sample mean of respondents in the Direct Report condition. Column 2 is the coefficient μ on "Veiled Report" from a regression with controls. Column 3 is the coefficient μ on "Veiled Report" when we used the same list items paired with a sensitive item of interest, rather than a placebo item of interest, estimated from a regression with controls. Standard errors in brackets: Columns 2 and 3 present heteroskedasticity-robust standard errors.

6.2 Inattentive Response

Can inattentive participants or "noise" explain our effects? First, note that the evidence is not consistent with any sizable portion of our participants simply responding randomly to our questions across both treatments. The easiest check of this is to refer to the histograms in Figure 1, which clearly document in two ways that most respondents are not simply randomly pressing buttons. First, the response distributions are not uniform. The average number of responses that are equal to zero or five (the boundaries) is 3.6%, substantially different than the 33% that would be present in a uniform distribution. Second, there are substantial differences across question in the way participants respond to our items. Third, the differences in treatment effects by demographics in ways predicted by existing evidence (i.e. Table 4) also provide evidence that participants are not responding randomly to the directly-asked sensitive item.

To our knowledge, there is no argument under which participants behave in the same manner across treatment (i.e. randomizing in both treatments) that would produce our pattern of treatment effects. Any potential confound story must rely on at least some subsample of our participants changing their behavior as a result of our treatment. We argue that the change in behavior can most reasonably be attributed to a lower cost of truth-telling in the veiled report method.

6.2.1 Robustness to Randomization of Responses

Nonetheless, suppose that some fraction of participants were "noise responders" who simply selected buttons at random. How would this affect the results? In the Veiled Report condition, participants faced a list of five statements and had to indicate how many of them were true. Noise responders would have $E[y_{qi}^V] = 2.5$, as they give 0, 1, 2, 3, 4 or 5 with equal probability.

In the Direct Report condition, participants faced a list of four statements and had to indicate how many of them were true. For noise responders, the expected number of these statements listed as true would be $E[c_{qi}] = 2$. When faced with the direct question about the sensitive answer, noise responders would pick true half the time, giving $E[d_{qi}] = 0.5$, and then the sum would be $E[y_{qi}^D] = 2.5$.

This type behavior would thus bias our differences across treatments toward zero, as $\mu_q = E[y_{qi}^V] - E[y_{qi}^D] = 0$ for noise responders. To the extent that individuals are responding inattentively in this way, our paper understates the treatment effect of the Veiled Report condition. For instance, if 10% of participants were noise responders, instead of a 7.3 percentage point increase in non-heterosexuality in Question 1, the increase would be 8.1 percentage points.

6.2.2 Robustness to Differential Randomization by Treatment, Worst-Case Analysis

There are many potential models of noise, and so in this section, we consider a worst-case analysis: what would be necessary to get our result if the Veiled Report method did not increase the rate of truth-telling at all, but instead led participants to respond in a completely uninformative manner. We take the starkest case: suppose some individuals are "differential randomizers" who are fully honest in the Direct Report treatment, but simply randomize their answer to the sensitive item in the Veiled Report treatment. (There is no evidence in our paper or the literature that the treatment would or could cause people to behave this way, but we consider it as a worst case for bias simply for illustrative purposes.)

We solve for the fraction of participants ρ_q who must be differential randomizers in order to produce these results. We find that not only are the number of differential randomizers often required to be high, they also vary substantially by question. Note that given the attenuation bias produced by noise responders, we would need even more differential randomizers if there were any noise responders.

The calculation for Question 1 is as follows. We assume that the true rate[26] of non-heterosexuality is the 11.3 percent rate reported under the direct report, giving $E[d_{qi}]$ =1-0.113=0.887. We observe the average response to the direct report list of four statements $E[c_{qi}] = 2.08$. Turning to the Veiled Report condition, "differential randomizers" respond in the same way as participants in the Direct Report condition to those four statements[27], but then randomize over the fifth sensitive item, answering "yes" to the fifth sensitive item with 50% probability. Since we observe the average response to the five-item veiled report list $E[y_{qi}^V]$=2.87, we can then solve for ρ_q:

$$E[y_{qi}^V] = E[c_{qi}] + \rho_q \left(\tfrac{1}{2} + 0\right) + (1 - \rho_q)E[d_{qi}]$$

$$E[y_{qi}^V] = 2.08 + \rho_q \left(\tfrac{1}{2} + 0\right) + (1 - \rho_q)(0.887)$$

Giving ρ =0.25. The first right-hand-side term in the expression above captures the common response of differential randomizers across treatment to the four statements on the direct report list. Then, fraction ρ randomize the response to the fifth question, while $1 - \rho$ give the true answer, which is non-heterosexuality only 11.3 percent of the time.

---

[26] Note that the following calculations do not depend upon this being the true rate of non-heterosexuality; they just depend on this being the reported rate of non-heterosexuality among non-randomizers in both treatments.

[27] This assumes that participants in the Direct and Veiled report either (i) both respond honestly to the first four list items, or (ii) both respond in the same non-informative way to the first four list items. We have no reason to believe that one of these assumptions would not hold. If simply changing the size of the list changes behavior, we pick this up by modeling a differential treatment of the fifth item in the Veiled Report condition.

Thus, in order to produce our estimated treatment effects for Question 1, we would need fully a quarter of the participants in the veiled report treatment to randomize their answer to the sensitive item in the list, not randomize for the rest of the list, and not randomize for the direct response sensitive item. In Table 7, we produce ρ for each question. Not surprisingly, questions for which we estimate relatively large treatment effects (1, 3, 5, and 6 in particular) require sizable fractions of our sample employing this worst-case randomization strategy. Perhaps the most convincing evidence that this is not what is going on is that in order to produce the results we see in the paper, the fraction of our sample using this randomization strategy would have to vary substantially across question. To generate the treatment effect for Question 1, we would need approximately 25 percent of our sample employing this randomization strategy; however, if 25 percent of our sample employed this strategy on Question 2, we would have observed a 9 percentage point treatment effect, well in excess of our actual observed effect. This type of randomization is also incompatible with the negative treatment effects observed in Questions 8, 9, and 12. Of course, the fact that our treatment effects vary with demographic information in intuitive ways also suggest that randomization is not driving our results.

A related model of inattentive response is one in which some subset of participants randomize choices when faced with a list, but respond honestly when asked a question directly. In this model, when faced with the five-item list in the Veiled Report condition, these participants would have $E\left[y_{qi}^{V}\right] = 2.5$. In the Direct Report condition, with the four-item list, these participants would have would have $E[c_{qi}] = 2$. Thus, for this subset of participants, we would estimate a treatment effect of $0.5 - E[d_{qi}]$ for each question with "yes" as the sensitive answer, and $(0.5 - (1 - E[d_{qi}]))$ for each question with "no" as the sensitive answer. We could ask whether our treatment effects could be produced by a population in which some fraction, r_q, used this strategy, while $(1 - r_q)$ responded honestly across all conditions, producing an estimated treatment effect of 0. That is, we could solve for the r_q such that:

$$E\left[y_{qi}^{V}\right] - E\left[y_{qi}^{D}\right] = r_q(0.5 - E[d_{qi}]) + (1 - r_q)(0)$$

This calculation actually reduces to the same calculation reported above for differential randomization, since $E\left[y_{qi}^{D}\right] = E[c_{qi}] + E[d_{qi}]$. Substituting and rearranging gives:

$$E\left[y_{qi}^{V}\right] = E\,c_{qi}\; r_q(0.5) + (1 - r_q)E\,d_{qi}$$

Table 7. Differential Randomization Needed to Generate Treatment Effects, Worst-Case Analysis

Question	Assumed Randomization Rate in Direct Report	Necessary Randomization Rate in Veiled Report (ρ)
Question 1 - Heterosexual	0	0.25
Question 2 - Attraction	0	0.003
Question 3 - Experience	0	0.36
Question 4 - Marriage	0	0.17
Question 5 - Manager	0	0.35
Question 6 - Discriminate	0	0.29
Question 7 - Adopt	0	0.18
Question 8 - Change	0	-0.26
Question 9 – Refuse to Serve	0	-0.14
Question 10 – Open at Work	0	-0.01
Question 11 – Co-worker	0	0.24
Question 12 – Work with Children	0	-0.29

Note: "Differential randomization" is precisely randomizing the answer to the sensitive item in the list but not randomizing the rest of the items in the list or the response to the direct sensitive item. Calculations assume zero respondents provide purely random responses. Including such respondents would increase the proportion of necessary differential randomizers.

6.3 Does the framing of the sensitive answer matter?

Our results do show an unexpected pattern. The ICT seems to work better at removing social desirability bias if the stigmatized response is "no". In the twelve sensitive questions studied (see Table 2 and Table 5), six had "no" as the stigmatized response, and six had "yes". All six "no" items produced significant results in the hypothesized direction. Only one of the "yes" items yielded significance in the anticipated direction. There are three reasonable explanations for this correlation. The first is that it is merely a chance relationship with only twelve observations. The second two we discuss in depth below.

The second possibility, and main concern, is that the observed treatment effects are a result of a mechanical downward bias in the ICT mechanism. Recall answering "no" to an item in the ICT means not adding one to your count, so an increase in no's would mean a negative treatment effect. If the ICT,

for any reason, produces a systematic negative treatment effect, this may enable researchers to find false treatment effects for sensitive items for which the sensitive answer is "no" (and to fail to find treatment effects for "yes" questions). We can answer this conjecture with our placebo data. As was noted in Section 6.1, only one of the eight placebo items produced a significant effect. The effect was indeed negative. Recall though that when the same list was tested with a different placebo item, no significant effect was found. Overall, the lists produced no consistent treatment effects even though the exact same lists were used when significant effects were found with sensitive items. We find no evidence the treatment effects found with the sensitive items are driven by a significant downward bias in the ICT.

The third possibility is that the ICT more effectively removes stigma for items whose sensitive answer is "no". For whatever reason, it may be that the ICT works at increasing honest responses, but works better, or only, if the respondent socially disapproved response is a "no". We will not be able to isolate any mechanism, but perhaps saying "yes" to a sensitive item and actively adding one to your running tally feels more stigmatized than saying "no" to a sensitive item and passively not adding to the tally.

To test whether the ICT works better for "no" responses, we ran a new study in which we tested sensitive questions in two ways: one where the sensitive answer is "yes," and the other where it is "no". To do this, we manipulated the language used in the items. We aimed to change the wording without actually changing what the question was asking. For example, the complement to "Do you consider yourself to be non-heterosexual?" was "Do you consider yourself to be heterosexual?" The list of items used, with exact wording, can be found in Table 8.

Study participants saw each of the four sensitive items, either in a list or directly, but only saw one version of each item; they did not see both an item and its complement. The study mirrored the previous two in every other way. It was run on Amazon Mechanical Turk, from July 20 to August 3, 2015, and 5,832 subjects participated (randomized across four treatments, Direct or Veiled Report, crossed with the framing of the sensitive answer).

As can be seen in Table 8, responses to all four sensitive items increased when the sensitive response was "no", achieving statistical significance twice. The same was not true when the sensitive response was "yes": Three decreased, one increased, and the only significant effect was a decrease. In regressions with controls (as in Equation 1), the average increase of sensitive "no's" was 6.3 percentage points (and 38.6% over the direct response estimates) while sensitive "yeses" actually decreased on average by 1.2 percentage points.

These results alongside the fact that the placebo studies find no significant negative bias, suggest the ICT is not producing false negative effects, but rather that the mechanism works better (or maybe

only) for "no" responses. Why this is true, and what this means for how surveying should be done, are interesting questions for future work.

Table 8. The ICT Works Better when "No" is the Sensitive Answer

Item Wording	Sensitive Answer	Direct Response Estimate	Increase under ICT
Do you consider yourself to be non-heterosexual?	Yes	14.0%	-0.1 [3.3]
Do you consider yourself to be heterosexual?	No	11.9%	10.1 [3.2]***
Have you had a sexual experience with a member of same sex?	Yes	20.6%	1.1 [3.9]
Have all your sexual experiences been with members of same sex?	No	22.8%	7.2 [3.7]**
Should it be legal to discriminate in hiring?	Yes	11.9%	-0.1 [3.1]
Should it be illegal to discriminate in hiring?	No	16.0%	4.3 [3.1]
Should be legal for small businesses to refuse to serve LGBT?	Yes	23.4%	-5.8%*
Should be illegal for small businesses to refuse to serve LGBT?	No	33.0%	3.7% [3.3]

Notes: n = 2,912 or 2,913 for "yes" items and n is between 2,914 and 2,917 for "no" items. Column 3 is the sample mean of respondents in the Direct Report condition. Column 4 is the coefficient μ on "Veiled Report" from a regression with controls (Equation 1). Robust standard errors in brackets.

7.   Discussion

The difference between our Direct Report and Veiled Report elicitations shows that current estimates—even though elicited under privacy and anonymity— do not correctly reflect true rates of non-heterosexual identity and same-sex sexual encounters, nor do they correctly gauge sentiment and political opinions on LGBT-related issues. That there is stigma attached to reporting anti-gay sentiments is perhaps quite surprising. All of the anti-gay positions considered in our five sentiment questions are either public

policy in many portions of the U.S., or have been advocated for by major political figures.[28] The fact that these opinions are still misrepresented suggests that many other opinions on controversial public issues may not be accurately measured.

Our findings provide insight into social norms surrounding sexuality. The decreased rate of reporting as heterosexual in the Veiled Report treatment suggests a societal stigma of being LGBT. At the same time, our data show that individuals are reluctant to report that they have attitudes or policy opinions that are not accepting of LGBT individuals, consistent with a stigma of holding *anti*-gay sentiments. Even though average sentiment in the United States has become more accepting of LGBT rights, we find that many LGBT individuals do not truthfully report their sexuality, even in a highly private and anonymous setting where the risks associated with truth-telling are arguably minimized. Thus, our data suggests that the stigma felt by many in this population has not been eliminated. This finding provides insights for a model of what is sufficient for stigma (e.g. can a small minority create a stigma for another group?).

As with any experiment with a non-representative sample, the next step is to think about how the results would generalize. Our experiment shows that the privacy afforded by standard survey methodology is not sufficient for honest responses to sensitive questions, specifically with respect to sexual identity and related attitudes. With appropriate qualifications, we can examine generalizability by examining selection into our survey. A benefit of our sample is that, while not representative of the U.S. as a whole, it has broad coverage of demographic characteristics that correlate with varying levels of misrepresentation.

Our sample is younger, more liberal, less religious, and slightly more Caucasian than the American adult population. The undersampled groups – older, conservative, religious, and from a minority group – all misrepresented their own sexual identity at a higher rate than our sample average (see Table 4). Though we cannot rule out that participants in our sample have different treatment effects due to unobservable factors, these interactions suggest that a larger treatment effect for sexual identity might be expected for a representative population.

For the anti-gay sentiments questions, however, the correlations between our undersampled groups and degrees of misrepresentation are less systematic. Though our theory predicts larger treatment effects for the sentiment questions for our oversampled groups, e.g. liberals, this was not conclusive in the data. As a result, it is more challenging to speculate in which direction, if any, the results may change for a representative sample. For one of the questions, support of same-sex marriage, there is now evidence

---

[28] For instance, as of this writing only 13 states issue licenses for same-sex marriages, and only 21 states prohibit employment discrimination based on sexual orientation. A number of leading political figures argue that homosexuality is "a choice," and adoption laws are in flux in many states and in some states explicitly ban same-sex couples from adopting.

that the treatment effects from our experiment are very similar to the treatment effects found using a representative sample. In a recent working paper begun after our own study, Lax, Phillips, and Stollwerk (2014), report the results of an ICT experiment on approval of same-sex marriage with a representative sample. Overall, similar to our data, they find an insignificant difference between estimated support of same-sex marriage under the list method and direct reporting. However, they do find a significant treatment effect (less support for same-sex marriage under veiled reporting) among Democrats, the same as in our sample. The similarity of treatment effects between their non-online, representative sample and our sample lends some confidence that the results we report here provide valuable insights into the population beyond Mechanical Turk. Representative samples aside, at a minimum, our data tell us that for a large and diverse group of U.S. residents, the level of privacy and anonymity provided by most current surveys on stigmatized opinions and behaviors is insufficient for eliciting honest answers from many respondents.

The misreporting of sexual identity and sexuality-related opinions that we observe has far-reaching implications. If LGBT identity is underreported, it suggests that other items related to that identity may also be underreported: for instance, data on workplace or housing discrimination, hate crimes, or domestic violence. Underreporting of this type may induce distortions in policies that rely on estimates of the size or characteristics of the LGBT population or the frequency of same-sex sex—for instance, workplace policies, the cost-benefit analysis of LGBT-related public health interventions, elder services, domestic violence prevention programs, and youth mental health/suicide prevention programs.

Privacy and anonymity, as in standard data-collection, are insufficient to get truthful reports about sensitive behaviors. Our results speak directly to LGBT-related issues, but should also lead managers and researchers to investigate how social desirability bias affects self-reported data in other domains.

Disclosure: We confirm that, for all experiments, we have reported all measures, conditions, data exclusions, and how we determined out sample sizes.

References

Ahmed, A. and Hammarstedt, M. 2010. "Sexual orientation and earnings: a register data-based approach to identify homosexuals". *Journal of Population Economics*, 23(3):835-549.

Akerlof, George A, and Rachel E. Kranton. "Economics and Identity." *The Quarterly Journal of Economics*, 115.3 (2000): 715-753.

Allegretto, Sylvia, and Michelle M. Arthur (2001). "An Empirical Analysis of Homosexual/Heterosexual Male Earnings Differentials: Unmarried and Unequal?" *Industrial and Labor Relations Review* 54: 631-646.

Ayres, Ian, and Peter Siegelman. "Race and gender discrimination in bargaining for a new car." *The American Economic Review* (1995): 304-321.

Badgett, M.V. Lee (2001). Money, Myths, and Change: The Economic Lives of Lesbians and Gay Men. Chicago: University of Chicago Press.

Banks, Christopher. The cost of homophobia: Literature review on the human impact of homophobia in Canada. Community-University Institute for Social Research, 2003.

Berg, N. and D. Lien. "Same-sex sexual behaviour: US frequency estimates from survey data with simultaneous misreporting and non-response," *Applied Economics* 38, 757–769 (2006).

Bertrand, Marianne, and Sendhil Mullainathan. "Implicit discrimination." *The American Economic Review* 95.2 (2005): 94-98.

Black, Dan, Hoda Makar, Seth Sanders, and Lowell Taylor (2003). "The Earnings Effects of Sexual Orientation." *Industrial and Labor Relations Review* 56: 449-469

Black, Dan, Gary Gates, Seth Sanders and Lowell Taylor. "Demographics of the gay and lesbian population in the United States: Evidence from available systematic data sources." *Demography* 37.2 (2000): 139-154.

Black, Dan, Gary Gates, Seth Sanders and Lowell Taylor. "Why Do Gay Men Live in San Francisco?." *Journal of Urban Economics* (2002): 54-76.

Black, Dan, Gary Gates, Seth Sanders and Lowell Taylor. "The Measurement of Same-Sex Unmarried Partner Couples in the 2000 U.S. Census." *California Center for Population Research Online Working Paper Series.* (2007) #CCPR-023-07.

Black, Dan, Seth Sanders and Lowell Taylor. "The Economics of Lesbian and Gay Families." *Journal of Economic Perspectives* 21(2) (2007): 53-70.

Blair, Graeme, Kosuke Imai and Jason Lyall. 2014. "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science* 58(4), 1043–1063.

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

Bloom, David E., and Sherry Glied. "Projecting the number of new AIDS cases in the United States." *International Journal of Forecasting* 8.3 (1992): 339-365.

Carpenter, Christopher. "Sexual Orientation and Outcomes in College." Economics of Education Review. (2007).

Chandra, A., W. Mosher, C. Copen, and C. Sionean. "Sexual Behavior, Sexual Attraction, and Sexual Identity in the United States: Data from the 2006-2008 National Survey of Family Growth." *National Health Statistics Reports* 36, (2011).

Clain, Suzanne H., and Karen Leppel (2001). "An Investigation into Sexual Orientation Discrimination as an Explanation for Wage Differences." *Applied Economics* 33: 37-47.

CNN ORC Poll. 2012. ORC International. Conducted May 29 – 31, 2012. Online: http://i2.cdn.turner.com/cnn/2012/images/06/06/rel5e.pdf

Coutts, Elisabeth, and Ben Jann. "Sensitive questions in online surveys: experimental results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)." *Sociological Methods & Research* 40.1 (2011): 169-193.

Dalton, Dan R., James C. Wimbush, and Catherine M. Daily. "Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior." *Personnel Psychology* 47.4 (1994): 817-829.

Das, Aniruddha, and EO Laumann. "How to Get Valid Answers from Survey Questions: What We Learned from Asking about Sexual Behavior and the Measurement of Sexuality." in *The Sage Handbook of Measurement,* G Walford, E Tucker Eds. (Sage Pubs, London 2010) p. 9-26.

Dasgupta, Nilanjana, and Luis M. Rivera. "From Automatic Antigay Prejudice to Behavior: The Moderating Role of Conscious Beliefs about Gender and Behavioral Control." Journal of Personality and Social Psychology 91 (2006): 268–280.

Edwards, Allen Louis. The social desirability variable in personality assessment and research. Dryden Press, 1957.

Ellison, Gavin, and Briony Gunstone. Sexual orientation explored: A study of identity, attraction, behaviour and attitudes in 2009. Equality and Human Rights Commission, 2009.

Fay, Robert, Charles Turner, Albert Klassen and John Gagnon. "Prevalence and patterns of same-gender sexual contact among men." *Science* 243.4889 (1989): 338-348.

Fershtman, Chaim, and Uri Gneezy. "Discrimination in a segmented society: An experimental approach." *The Quarterly Journal of Economics* 116.1 (2001): 351-377.

Fisher, Robert J. "Social desirability bias and the validity of indirect questioning." *Journal of Consumer Research* (1993): 303-315.

Gates, Gary. "How Many People are Lesbian, Gay, Bisexual, and Transgender?" *Williams Institute Report* (2011).

Glover D, Pallais A, Pariente W. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." Working Paper. (2015).

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. "Understanding and Using the Implicit Association Test: Meta-Analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97 (2006): 17–41.

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.

Herek, Gregory M., and Eric K. Glunt. "Interpersonal contact and heterosexuals' attitudes toward gay men: Results from a national survey." *Journal of Sex Research* 30.3 (1993): 239-244.

Holbrook, Allyson L., and Jon A. Krosnick. "Social desirability bias in voter turnout reports Tests using the item count technique." *Public Opinion Quarterly* 74.1 (2010): 37-67.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. "The online laboratory: Conducting experiments in a real labor market." *Experimental Economics* 14.3 (2011): 399-425.

Jamison, Julian C., Dean Karlan, and Pia Raffler. "Mixed-Method Evaluation of a Passive mHealth Sexual Information Texting Service in Uganda." Information Technologies and International Development 9(3) (2013).

Jepsen, Lisa K. (2007). "Comparing the Earnings of Cohabiting Lesbians, Cohabiting Heterosexual Women, and Married Women: Evidence from the 2000 Census." *Industrial Relations.*

Jepsen, Lisa and Christopher Jepsen. "An Empirical Analysis of the Matching Patterns of Same-Sex and Opposite-Sex Couples." *Demography* 39 (2002): 435-453.

John, L, Loewenstein G., Acquisti A., Vosgerai J. "Paradoxical Effects of Randomized Response Techniques". In submission. *Journal of Marketing Research.* 2013.

Kinsey, Alfred, Wardell Pomeroy, and Clyde Martin. *Sexual Behavior in the Human Male*. Indiana University Press (1948).

Klawitter, Marieka and Victor Flatt. "The effects of state and local antidiscrimination policies on earnings for gays and lesbians." *Journal of Policy Analysis and Management* 17.4 (1998):658-686.

Kuran, Timur. Private Truths, Public Lies: The Social Consequences of Preference Falsification. Harvard University Press (1995).

LaBrie, Joseph W., and Mitchell Earleywine. "Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique." *Journal of Sex Research* 37.4 (2000): 321-326.

Lax, Jeffrey R., Justin H. Phillips, and Alissa F. Stollwerk. "Are Respondents Lying About their Support of Same-Sex Marriage? Lessons from A Recent List Experiment." Working paper. 2014.

Lensvelt-Mulders, G., J. Hox, and P. van der Heijden. "How to Improve the Efficiency of Randomised Response Designs" *Quality and Quantity* 39.5 (2005): 253-265.

Li, F., and V. Nagar. 2013. Diversity and performance. *Management Science* 59 (3): 529-544.

Lundberg, Shelly, and Robert A. Pollak. 2007. "The American Family and Family Economics" *Journal of Economic Perspectives*, 21(2): 3-26.

Maccoby, E. E. and N. Maccoby. "The Interview: A tool of social science." In G. Lindzey (Ed.), *Handbook of Social Psychology: Vol. 1 Theory and Method.* Cambridge, MA: Addison Wesley: 449-487.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. "Temporal Distance and Discrimination An Audit Study in Academia." *Psychological Science* 23.7 (2012): 710-717.

Miller, JD. *A new survey technique for studying deviant behavior*. (PhD Diss. G Wash U, 1984)

Miller, K. 2001. "Cognitive Testing of the NHANES Sexual Orientation Questions." Report for the Office of Research on Women's Health, National Institutes of Health.

Office of National Statistics, "Developing Survey Questions on Sexual Identity: Rationale and Design of Sexual Identity Questioning on the Integrated Household Survey" (2008)

Oreffice, Sonia. "Sexual orientation and household decision making: Same-sex couples' balance of power and labor supply choices." *Labour Economics* 18.2 (2011): 145-158.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5.5 (2010): 411-419.

Pew Research Report. "A Survey of LGBT Americans" June 2013.
Online: http://www.pewsocialtrends.org/2013/06/13/a-survey-of-lgbt-americans/

Powell, Richard. "Social Desirability Bias in Polling on Same-sex Marriage Ballot Measures." *American Politics Research* 41 (2013): 1052-1070.

Rand, David G. "The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments." *Journal of Theoretical Biology* 299 (2012): 172-179.

Rayburn, Nadine Recker, Mitchell Earleywine, and Gerald C. Davison. "An investigation of base rates of anti-gay hate crimes using the unmatched-count technique." *Journal of Aggression, Maltreatment & Trauma* 6.2 (2003): 137-152.

Stephens-Davidowitz, Seth. "The Cost of Racial Animus on a Black Presidential Candidate: Using Google Search Data to Find What Surveys Miss." SSRN Working Paper (2013).

Tourangeau, Roger, and Ting Yan. "Sensitive questions in surveys." *Psychological Bulletin* 133.5 (2007): 859.

Tsuchiya, Takahiro, Yoko Hirai, and Shigeru Ono. "A study of the properties of the item count technique." *Public Opinion Quarterly* 71.2 (2007): 253 – 272.

Zimmerman, Rick S., and Lilly M. Langer. "Improving estimates of prevalence rates of sensitive behaviors: The randomized lists technique and consideration of self-reported honesty." *Journal of Sex Research* 32.2 (1995): 107-117.

Appendix: For Online Publication

Table A1. Order Effects

		Order	
		Forward	Reverse
	Direct Report		
	Mean Reported Number of True Four Items	2.07	2.09
		[0.85]	[0.90]
	Percent Reporting Sensitive Answer	10 percent	13 percent
		[1.19]	[1.32]
Question 1	N	631	639
	Veiled Report		
	Mean Reported Number of True Items	2.91	2.84
		[0.99]	[0.99]
	N	618	628
	Direct Report		
	Mean Reported Number of True Four Items	2.26	2.21
		[0.84]	[0.81]
	Percent Reporting Sensitive Answer	25 percent	20 percent
		[1.71]	[1.58]
Question 8	N	631	639
	Veiled Report		
	Mean Reported Number of True Items	2.4	2.39
		[0.96]	[0.88]
	N	618	628

Table A2. Estimates for Full Sample & Respondents Reporting Consistent Geographic Data

Sensitive Answer	Percent Reporting Sensitive Answer Under Direct Report		Change in Reporting of Sensitive Answer Under Veiled Report	
	Full Sample	Respondents Reporting Consistent Zip and State	Full Sample	Respondents Reporting Consistent Zip and State
Own Sexuality				
Not Heterosexual	11.3	10.8	7.3	7.7
	[0.89]	[1.03]	[3.57]	[4.14]
Same sex Attraction	13.9	13.8	1.3	2.8
	[0.97]	[1.14]	[3.63]	[4.21]
Same-sex sexual experience	17.2	16.2	10.1	12.4
	[1.06]	[1.22]	[3.82]	[4.41]
Attitudes and Opinions about Issues Related to Sexual Orientation				
Not Support Same-sex Marriage	18.8	19.5	4.2	0.7
	[1.10]	[1.31]	[3.18]	[3.85]
Not Happy with LGB Manager	16.2	17.0	10.8	16.4
	[1.03]	[1.24]	[3.75]	[4.62]
Not Illegal to Discriminate	14.4	13.2	10.3	11.5
	[0.99]	[1.12]	[3.33]	[4.08]
LGB not allowed to Adopt	12.9	13.1	5.9	7.6
	[0.94]	[1.12]	[3.40]	[3.80]
Can Change Orientation	22.2	24.0	-7.0	-5.8
	[1.17]	[1.41]	[3.52]	[4.33]

*n= 1793, with 916 in Direct Ask condition. Columns 1 and 2 are sample means. Columns 3 and 4 are the coefficients μ on "Veiled Report" from a regression with controls. In Columns 3 and 4, standard errors are derived using the bootstrap.

Table A3. Direct Report Design Differences Across Wave

First Wave Direct Report	Second Wave Direct Report
• I remember where I was the day of the Challenger space shuttle disaster.	• I remember where I was the day of the Challenger space shuttle disaster.
• I spent a lot of time playing video games as a kid.	• I spent a lot of time playing video games as a kid.
• I would vote to legalize marijuana if there was a ballot question in my state.	• I would vote to legalize marijuana if there was a ballot question in my state.
Please fill in the bubble that corresponds to the total number of statements above that apply to you.	• I have voted for a political candidate who is pro-life.
0 1 2 3	**Please fill in the bubble that corresponds to the total number of statements above that apply to you.**
Have you voted for a political candidate who is pro-life?	0 1 2 3 4
Yes No	
Do you consider yourself to be heterosexual?	**Do you consider yourself to be heterosexual?**
Yes No	Yes No

Table A4. Estimated Treatment Effects by Wave of Data Collection

Sensitive Answer	ΔReporting of Sensitive Answer under Veiled Report		
	First Wave	Second Wave	Overall
Not Heterosexual	8.9	6.2	7.3
	[6.82]	[4.41]	[3.57]
Same sex Attraction	13.1	-4.6	1.3
	[6.61]	[4.49]	[3.63]
Same-sex sexual experience	8.2	9.8	10.1
	[6.92]	[4.76]	[3.82]
Not Support Same-sex Marriage	4.3	4.3	4.2
	[5.71]	[3.99]	[3.18]
Not Happy with LGB Manager	16.8	8	10.8
	[6.58]	[4.72]	[3.75]
Not Illegal to Discriminate	5.9	13.3	10.3
	[6.24]	[4.09]	[3.33]
LGB not allowed to Adopt	9.3	6.4	5.9
	[6.13]	[4.27]	[3.40]
Can Change Orientation	-9.3	-6.4	-7
	[6.24]	[4.41]	[3.52]

Table A5. Detailed Descriptive Statistics

	Direct Report	Veiled Report	Overall
Age			
30 and under	66 percent	66 percent	66 percent
31-50	28 percent	28 percent	28 percent
over 50	6 percent	6 percent	6 percent
Median age	26.5	26	26
Mean age	30	30	30
			(10.44 SD)
Gender			
Male	58 percent	57 percent	57 percent
Female	41 percent	43 percent	42 percent
Transgender	0.4 percent	0.7 percent	0.6 percent
Race			
White	80 percent	80 percent	80 percent
Black	6 percent	6 percent	6 percent
Hispanic*	6 percent	8 percent	7 percent
Asian	8 percent	7 percent	7 percent
Indian	2 percent	1 percent	2 percent
Relationship Status			
Single	44 percent	41 percent	43 percent
Unmarried but in a relationship	23 percent	27 percent	25 percent
Married	28 percent	29 percent	28 percent
Divorced	5 percent	4 percent	4 percent
Parental Status			
Parent	29 percent	29 percent	29 percent
Employment Status			
Works Full Time	39 percent	38 percent	38 percent
Works Part Time	17 percent	19 percent	18 percent
Unemployed	14 percent	15 percent	14 percent
Chooses to stay at home	5 percent	7 percent	6 percent
Retired	3 percent	1 percent	2 percent
Current Student	23 percent	21 percent	22 percent
Region			
Northeast	20 percent	19 percent	20 percent
Midwest	22 percent	23 percent	22 percent
South	33 percent	35 percent	34 percent

West	25 percent	24 percent	24 percent
Education			
Some High School	1 percent	2 percent	1 percent
Finished High School	11 percent	11 percent	11 percent
Some College	44 percent	43 percent	43 percent
Finished College	30 percent	32 percent	31 percent
Some Graduate School	5 percent	5 percent	5 percent
Finished Graduate School	9 percent	8 percent	9 percent
Religion			
Christian	37 percent	36 percent	36 percent
Jewish	2 percent	2 percent	2 percent
No Religion	42 percent	44 percent	43 percent
Median response to "how religious are you" on 1-7 scale	2	2	2
Mean response to "how religious are you" on 1-7 scale	2.66	2.68	2.67 (2.00 SD)
Political Views			
Republican	15 percent	17 percent	16 percent
Democrat	46 percent	46 percent	46 percent
Independent	33 percent	31 percent	32 percent
Median response to "how political are you" on 1-7 scale	5	5	5
Mean response to "how political are you" on 1-7 scale	4.40	4.33	4.37 (1.73 SD)
Answered the Cognitive Reflection Task Correctly	39 percent	40 percent	40 percent

*Note: estimate for Hispanic comes from the 1730 participants of whom this question was asked.

Table A6. Comparisons of Means Across Treatment

	Mean Total # of Items in Direct Report, $\bar{y}_q^D$	Mean Total # of Items in Veiled Report, $\bar{y}_q^V$	ΔReporting of Sensitive Answer from Mean Comparison	ΔReporting of Sensitive Answer from Regression Table 3
Question 1 - Heterosexual	2.97 [0.930]	2.87 [0.992]	0.096 [0.038]	0.073 [0.036]
Question 2 - Attraction	2.20 [0.902]	2.20 [0.962]	0.005 [0.019]	0.013 [0.036]
Question 3 - Experience	1.92 [0.985]	2.04 [0.978]	0.116 [0.039]	0.101 [0.038]
Question 4 - Marriage	2.65 [0.867]	2.60 [0.870]	0.054 [0.035]	0.042 [0.032]
Question 5 - Manager	3.17 [0.937]	3.05 [0.990]	0.120 [0.038]	0.108 [0.038]
Question 6 - Discriminate	2.52 [0.838]	2.41 [0.843]	0.107 [0.033]	0.103 [0.033]
Question 7 - Adopt	1.92 [0.872]	1.86 [0.894]	0.063 [0.035]	0.059 [0.034]
Question 8 - Change	2.46 [0.915]	2.39 [0.921]	-0.067 [0.037]	-0.070 [0.035]

Notes: *n= 2516, with 1270 in Direct Report condition. Column 1 is the mean number of "Yes" answers reported in the Direct Report condition (summed over the sensitive item and the number of N items). Column 2 is the mean number of "Yes" answers reported in the Veiled Report condition (the number of N+1 items the respondent agreed to). Column 3 is the difference between Column 1 and Column 2, with the sign adjusted to give the change in number of sensitive answers. Column 4 gives the corresponding regression-adjusted estimate. Columns 1 and 2: Standard deviations in brackets. Columns 3 and 4: standard errors in brackets.

Table A7. Number of Sensitive Answers Provided under Direct Report for Own Sexuality Questions, by Demographics

	1-Hetero.	2-Attract.	3-Experience	N		1-Hetero.	2-Attract.	3-Experience	N
Gender					*Politics*				
Male	0.0797	0.0919	0.124	740	Democrat	0.149	0.182	0.208	578
	[0.00988]	[0.0106]	[0.0123]			[0.0152]	[0.0155]	[0.0171]	
Female	0.160	0.206	0.238	525	Republican	0.0515	0.0825	0.124	194
	[0.0161]	[0.0182]	[0.0189]			[0.0161]	[0.0194]	[0.0236]	
Race					*Age*				
White	0.115	0.144	0.188	1022	Under 31	0.131	0.161	0.173	840
	[0.00962]	[0.0108]	[0.0123]			[0.0116]	[0.0130]	[0.0130]	
Black	0.0864	0.136	0.0988	81	31 – 50	0.0855	0.111	0.182	351
	[0.0310]	[0.0386]	[0.0329]			[0.0144]	[0.0167]	[0.0205]	
Religion					51 plus	0.0506	0.0380	0.127	79
Christian	0.0821	0.0994	0.112	463		[0.0246]	[0.0190]	[0.0396]	
	[0.0132]	[0.0137]	[0.0152]						
No Religion	0.135	0.164	0.230	535					
	[0.0152]	[0.0164]	[0.0186]						

Table A8. Number of Sensitive Answers Provided under Direct Report for Sentiment Questions, by Demographics

	4 - Marriage	5 - Manager	6 - Discriminate	7 - Adopt	8- Change	N
Gender						
Male	0.184	0.182	0.158	0.131	0.232	740
	[0.0144]	[0.0143]	[0.0138]	[0.0128]	[0.0156]	
Female	0.196	0.133	0.126	0.128	0.208	525
	[0.0174]	[0.0145]	[0.0142]	[0.0146]	[0.0172]	
Race						
White	0.172	0.142	0.133	0.115	0.199	1022
	[0.0115]	[0.0110]	[0.0106]	[0.0101]	[0.0126]	
Black	0.407	0.333	0.235	0.321	0.457	81
	[0.0538]	[0.0502]	[0.0454]	[0.0522]	[0.0550]	
Religion						
Christian	0.343	0.272	0.205	0.210	0.326	463
	[0.0220]	[0.0206]	[0.0187]	[0.0191]	[0.0225]	
No Religion	0.0355	0.0710	0.107	0.0355	0.110	535
	[0.00779]	[0.0113]	[0.0131]	[0.00806]	[0.0138]	
Politics						
Democrat	0.0969	0.104	0.106	0.0640	0.147	578
	[0.0128]	[0.0127]	[0.0130]	[0.00989]	[0.0141]	
Republican	0.479	0.345	0.227	0.309	0.392	194
	[0.0360]	[0.0351]	[0.0298]	[0.0326]	[0.0353]	
Age						
Under 31	0.150	0.148	0.126	0.0988	0.214	840
	[0.0127]	[0.0125]	[0.0116]	[0.0102]	[0.0141]	
31 – 50	0.236	0.160	0.157	0.151	0.228	351
	[0.0221]	[0.0193]	[0.0197]	[0.0187]	[0.0219]	
51 plus	0.380	0.329	0.278	0.354	0.278	79
	[0.0557]	[0.0521]	[0.0512]	[0.0553]	[0.0529]	

Table A9. Change in Sensitive Answer Reports for Sentiment Questions in Veiled Report, by Demographics

	4 - Marriage	5 – Manager	6 - Discriminate	7 - Adopt	8- Change	N
Gender						
Male	1.67	4.78	5.35	8.13	-2.28	1444
	[3.90]	[4.86]	[4.5]	[4.41]	[4.58]	
Female	5.83	18.2	16.2	3.07	-10.8	1058
	[5.33]	[5.93]	[5.02]	[5.44]	[5.57]	
Race						
White	4.71	11.1	10.5	7.86	-6.56	2022
	[3.55]	[4.17]	[3.68]	[3.72]	[3.92]	
Black	6.75	15.7	28.1	9.8	-3.02	151
	[14.2]	[16.8]	[13.9]	[18.1]	[17]	
Religion						
Christian	3.7	12.7	17.5	10.0	-9.35	905
	[6.08]	[6.34]	[5.67]	[6.02]	[6.23]	
No Religion	3.62	10.8	6.8	7.48	-0.317	1078
	[4.09]	[5.57]	[4.98]	[4.74]	[5.16]	
Politics						
Democrat	9.71	14.3	9.06	3.58	-1.56	1155
	[4.28]	[5.3]	[4.75]	[4.9]	[5.06]	
Republican	5.69	32.1	23.9	16.9	-12.0	400
	[10.1]	[10.2]	[9.19]	[9.17]	[9.77]	
Age						
Under 31	4.98	11.7	7.35	8.98	-4.99	1658
	[3.76]	[4.6]	[4.24]	[4.23]	[4.12]	
31 – 50	1.03	11.6	14.2	1.24	-13.6	700
	[6.51]	[7.19]	[5.97]	[6.38]	[7.05]	
51 plus	17.7	16.5	13.4	11.8	-1.80	158
	[14.4]	[17.4]	[13.5]	[15.1]	[18.0]	

Note: n= 2516. The table gives the coefficient μ on "Veiled Report" from a regression with controls run on each demographic subgroup. (Equivalent to Column 2 of Table 3.) Heteroskedasticity-robust standard errors in brackets.

Appendix Section on Direct Question Only Treatment

One goal in designing our Direct Report treatment was to collect a baseline estimated prevalence of the sensitive answer using a directly asked question (i.e., "Do you consider yourself to be heterosexual?"). This would allow us to make comparisons of estimated prevalence under our method to the estimated prevalence under a common survey method, directly asked sensitive questions. We chose to include the list of four non-target items directly prior to the directly asked sensitive item in the Direct Report treatment to maintain as much parallelism as possible with our Veiled Report treatment. That way, if it is the case that something about the non-target items influences the way that the participant responds to the target item, then this force is present across both treatments and does not confound our estimates.

However, it does raise the possibility that our estimates from the directly asked questions do not reflect how participants would respond to the same directly asked question absent the non-target list on the same page. We chose to test this directly with a new treatment in Experiment 2. We randomized 100 participants from Experiment 2 into a "Direct Question Only" treatment; 88 completed the survey. These participants completed the same demographic survey the rest of participants completed, then proceeded to a modified version of the Direct Report treatment. In the Direct Question Only treatment, participants saw the eight directly asked questions of Experiment 2, but no non-target lists. That is, each page consisted of just one, directly asked question. Four were the placebo items (placebo items 1 – 4 discussed in Section 6.1), and four were sensitive items (sensitive items 9 – 12 discussed in Section 5.4).

In Table A10, we compare the responses of participants in the Direct Question Only treatment to participants in our standard Direct Report treatment. The fraction of "Yes" responses does not vary significantly across the two treatments.

Table A10 Evaluating the Direct Question Only Treatment

Question	Fraction "Yes" in Direct Question ONLY Treatment	Fraction "Yes" in Direct Report Treatment	Estimated Difference from Regression with Controls	p-value
Placebo – 1	0.612	0.531	-0.0818	0.12
Placebo – 2	0.316	0.268	0.0482	0.32
Placebo – 3	0.449	0.526	-0.0772	0.14
Placebo – 4	0.092	0.102	-0.0105	0.73
Sensitive - 9	0.184	0.198	-0.0147	0.72
Sensitive - 10	0.051	0.094	-0.0429	0.07
Sensitive - 11	0.898	0.903	-0.0046	0.88
Sensitive - 12	0.163	0.18	-0.017	0.66

Notes: *n=1533, with 98 in Direct Only Treatment. Column 1 is the fraction of "Yes" answers reported in the Direct Question Only condition. Column 2 is the fraction of "Yes" answers reported in the regular Direct Report condition. Column 3 gives the regression-adjusted estimate of the difference in fraction of "Yeses" reported in the Direct Question Only condition relative to the Direct Only Treatment in Experiment 2, estimated using controls. Column 4 reports the p-value on the coefficient on Direct Question Only condition from the regression with controls. Heteroskedasticity-robust standard errors used.
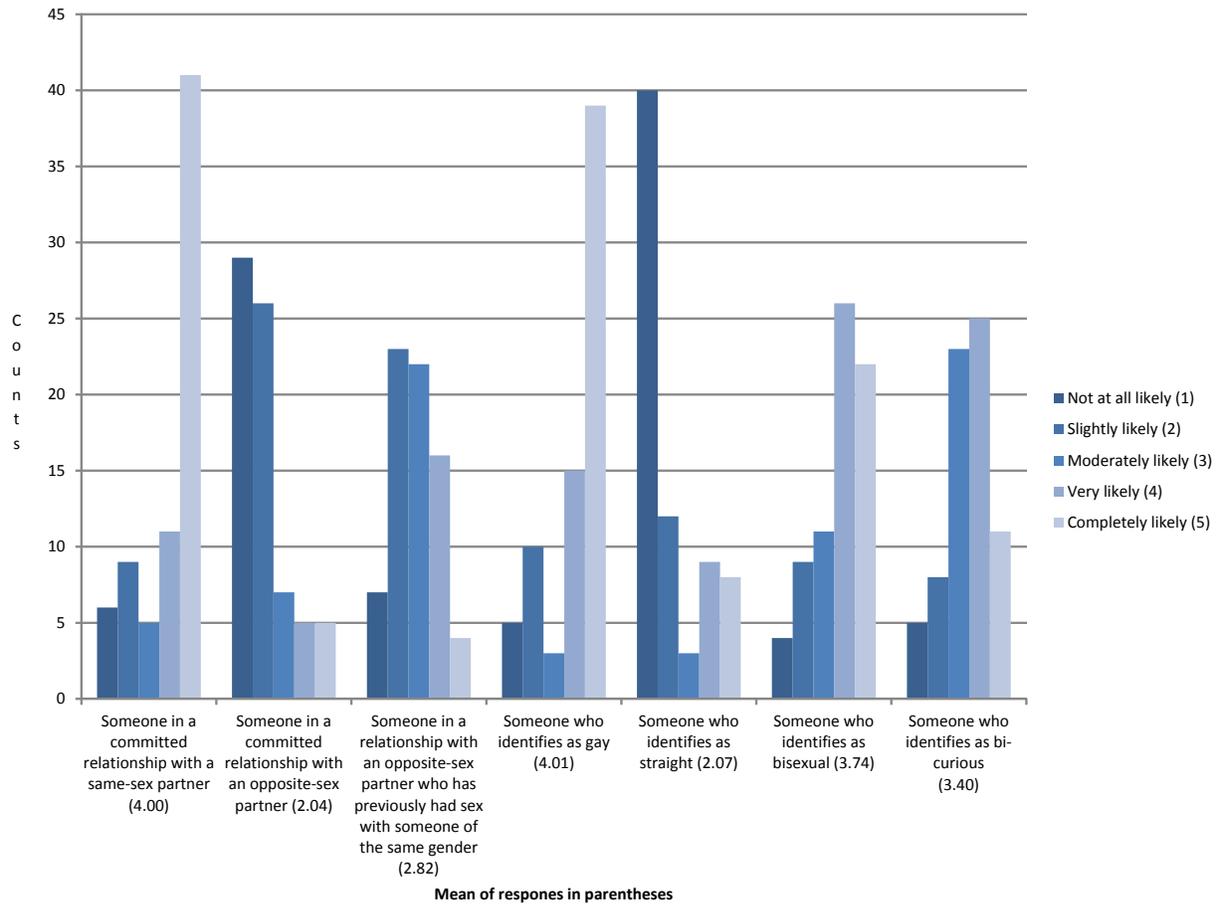
Figure A1. How likely is it that the following people would answer YES to the question, "Are you sexually attracted to members of the same sex?"

Materials

Below, we present a list of all questions used. We present the Direct Report questions as used in the second wave of data collection; see Table A4 for a comparison of the two waves. The question of whether the individual considered themselves to be of Hispanic, Latino, or Spanish origin was added in the second wave to supplement our data on race from the demographics section.

Demographics	
Gender	With what gender do you most closely identify?
Age	What is your age?
Race	What is your race? Select the race that best describes you.
Relationship	What best describes your current relationship status?
Parent	Are you a parent?
Employment	Employment Status:
State	In what state do you currently reside?
Zip Code from Youth	What is the zip code of the place you grew up? If you don't remember, you can write in the city and state.
Education	What is the highest level of education you've completed?
Religion Scale	On a scale of 1-7, 1 being not at all and 7 being extremely, how religious would you say you are?
Political Scale	On a scale of 1-7, 1 being not at all and 7 being extremely, how interested in politics would you say you are?
Religion	How would you describe your religious beliefs?
Politics	How would you describe your political affiliation?

Own Sexuality		
	Veiled Report	Direct Report
1 - Heterosexual	I remember where I was the day of the Challenger space shuttle disaster.	I remember where I was the day of the Challenger space shuttle disaster.
	I spent a lot of time playing video games as a kid.	I spent a lot of time playing video games as a kid.
	I would vote to legalize marijuana if there was a ballot question in my state.	I would vote to legalize marijuana if there was a ballot question in my state.
	I have voted for a political candidate who is pro-life.	I have voted for a political candidate who is pro-life.
	I consider myself to be heterosexual.	Please fill in the bubble that corresponds to the total number of statements above that apply to you.
	Please fill in the bubble that corresponds to the total number of statements above that apply to you.	Do you consider yourself to be heterosexual?
2 – Attraction	I get my town's newspaper delivered to my home.	I get my town's newspaper delivered to my home.
	I have applied for a job via the internet.	I have applied for a job via the internet.
	I think military spending should be cut.	I think military spending should be cut.
	I oppose stricter gun laws in this country.	I oppose stricter gun laws in this country.
	I am sexually attracted to members of the same sex.	Please fill in the bubble that corresponds to the total number of statements above that apply to you.
	Please fill in the bubble that corresponds to the total number of statements above that apply to you.	Are you sexually attracted to members of the same sex?
3 – Experience	I never or almost never buy organic food.	I never or almost never buy organic food.
	I drive a hybrid car or have considered getting one.	I drive a hybrid car or have considered getting one.
	I currently take a prescription medication.	I currently take a prescription medication.
	I own an iPod or other portable mp3 music player.	I own an iPod or other portable mp3 music player.
	I have had a sexual experience with someone of the same sex.	Please fill in the bubble that corresponds to the total number of statements above that apply to you.
	Please fill in the bubble that corresponds to the total number of statements above that apply to you.	Have you had a sexual experience with someone of the same sex?

Opinions about Sexuality

	Veiled Report	Direct Report
4 – Marriage	I attended church this Sunday. I am taller than 5 ft 6 inches. I have a sewing machine. I think Bill Clinton was a better president than George W. Bush. I think marriages between gay and lesbian couples should be recognized by the law as valid, with the same rights as heterosexual marriages. Please fill in the bubble that corresponds to the total number of statements above that apply to you.	I attended church this Sunday. I am taller than 5 ft 6 inches. I have a sewing machine. I think Bill Clinton was a better president than George W. Bush. Please fill in the bubble that corresponds to the total number of statements above that apply to you. Do you think marriages between gay and lesbian couples should be recognized by the law as valid, with the same rights as heterosexual marriages?
5 – Manager	I have more than one bathroom in my home. I can walk to a grocery store from my home in less than 30 minutes. I can speak more than one language. All four of my grandparents were born in the United States. I would be happy to have an openly lesbian gay or bisexual manager at work. Please fill in the bubble that corresponds to the total number of statements above that apply to you.	I have more than one bathroom in my home. I can walk to a grocery store from my home in less than 30 minutes. I can speak more than one language. All four of my grandparents were born in the United States. Please fill in the bubble that corresponds to the total number of statements above that apply to you. Would be happy to have an openly lesbian gay or bisexual manager at work?
6 - Discriminate	I am satisfied with my current financial situation. I expect the US will be in a world war within the next ten years. I have a driver's license. I often walk to work. I believe it should be illegal to discriminate in hiring based on someone's sexual orientation. Please fill in the bubble that corresponds to the total number of statements above that apply to you.	I am satisfied with my current financial situation. I expect the US will be in a world war within the next ten years. I have a driver's license. I often walk to work. Please fill in the bubble that corresponds to the total number of statements above that apply to you. Do you believe it should be illegal to discriminate in hiring based on someone's sexual orientation?

	I am an only child.	I am an only child.
	I live in the same state in which I was born.	I live in the same state in which I was born.
	It has been 10 or more years since I voted in a political election.	It has been 10 or more years since I voted in a political election.
7 – Adopt	I've worn a campaign button, put a campaign sticker on my car, or placed a sign in my window or in front of my house for a political candidate.	I've worn a campaign button, put a campaign sticker on my car, or placed a sign in my window or in front of my house for a political candidate.
	I believe lesbians and gay men should be allowed to adopt children.	Please fill in the bubble that corresponds to the total number of statements above that apply to you.
	Please fill in the bubble that corresponds to the total number of statements above that apply to you.	Do you believe lesbians and gay men should be allowed to adopt children?
	I have more than one email address.	I have more than one email address.
	I currently have a pet animal.	I currently have a pet animal.
	I have seasonal allergies.	I have seasonal allergies.
	I remember where I was when Reagan was shot.	I remember where I was when Reagan was shot.
8 – Change	I think someone who is homosexual can change their sexual orientation if they choose to do so.	Please fill in the bubble that corresponds to the total number of statements above that apply to you.
	Please fill in the bubble that corresponds to the total number of statements above that apply to you.	Do you think someone who is homosexual can change their sexual orientation if they choose to do so?

Post-Survey

Cognitive Reflection Task	Please do your best to answer the following question correctly. (1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
Risk	If you had a choice between the following two options, which would you prefer? (A) $20 for certain, (B) 50 percent chance of $0, percent50 chance of $60
Zip Code	What is the zip code where you currently live?
Hispanic	Do you consider yourself to be of Hispanic, Latino, or Spanish origin?
Confusion	Was there anything unclear or confusing about this experiment?