

NBER WORKING PAPER SERIES

THE SIZE OF THE LGBT POPULATION AND THE MAGNITUDE OF ANTI-GAY  
SENTIMENT ARE SUBSTANTIALLY UNDERESTIMATED

Katherine B. Coffman  
Lucas C. Coffman  
Keith M. Marzilli Ericson

Working Paper 19508  
<http://www.nber.org/papers/w19508>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2013

The authors acknowledge funding from Boston University and The Ohio State University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Katherine B. Coffman, Lucas C. Coffman, and Keith M. Marzilli Ericson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Katherine B. Coffman, Lucas C. Coffman, and Keith M. Marzilli Ericson

NBER Working Paper No. 19508

October 2013

JEL No. C90,D10,J10

**ABSTRACT**

Measuring sexual orientation, behavior, and related opinions is difficult because responses are biased towards socially acceptable answers. We test whether measurements are biased even when responses are private and anonymous and use our results to identify sexuality-related norms and how they vary. We run an experiment on 2,516 U.S. participants. Participants were randomly assigned to either a “best practices method” that was computer-based and provides privacy and anonymity, or to a “veiled elicitation method” that further conceals individual responses. Answers in the veiled method preclude inference about any particular individual, but can be used to accurately estimate statistics about the population. Comparing the two methods shows sexuality-related questions receive biased responses even under current best practices, and, for many questions, the bias is substantial. The veiled method increased self-reports of non-heterosexual identity by 65% ( $p < 0.05$ ) and same-sex sexual experiences by 59% ( $p < 0.01$ ). The veiled method also increased the rates of anti-gay sentiment. Respondents were 67% more likely to express disapproval of an openly gay manager at work ( $p < 0.01$ ) and 71% more likely to say it is okay to discriminate against lesbian, gay, or bisexual individuals ( $p < 0.01$ ). The results show non-heterosexuality and anti-gay sentiment are substantially underestimated in existing surveys, and the privacy afforded by current best practices is not always sufficient to eliminate bias. Finally, our results identify two social norms: it is perceived as socially undesirable both to be open about being gay, and to be unaccepting of gay individuals.

Katherine B. Coffman  
Department of Economics  
Ohio State University  
Arps Hall, 1945 N. High St  
Columbus, OH 43210  
baldiga.1@osu.edu

Keith M. Marzilli Ericson  
Boston University School of Management  
595 Commonwealth Avenue  
Boston, MA 02215  
and NBER  
kericson@bu.edu

Lucas C. Coffman  
Department of Economics  
Ohio State University  
Arps Hall, 1945 N. High St  
Columbus, OH 43210  
coffman.155@osu.edu

## Introduction

Accurately measuring sexual orientation and behaviors is important for policy, but difficult to do, as this topic is sensitive and perceived social stigma may drive biased responses. Research in a variety of areas uses data about the population that identifies as lesbian, gay, bisexual, or transgender (LGBT) or engages in same-sex sexual activity.<sup>1</sup> For instance, these data are relied upon by labor economists for studies of discrimination,<sup>2</sup> by urban economists for analyses of urban amenities,<sup>3</sup> and by health economists and public health researchers for estimating the prevalence of and optimal response to sexually-transmitted diseases.<sup>4</sup> Data on LGBT households has been used to test theories of the economics of the family, including household labor supply, educational investment, the demand for children, and the gender-based divisions of labor.<sup>5</sup> Moreover, the size of the LGBT population informs the prioritization of resource allocation. Many other fields also use estimates of the size of the LGBT population, including geneticists, evolutionary biologists, and psychologists (Gavrilets and Rice 2006, Cochran, Sullivan, and Mays 2003).

Data on LGBT-related sentiment, including opinions and beliefs regarding LGBT individuals and LGBT-related policy, can be used to inform analyses of discrimination and social disparities, as well as to inform the decisions of policy-makers.<sup>6</sup> Yet, these questions might also be sensitive. Survey answers might be biased towards social norms, suggesting that the widely-used data from polls and surveys may not be accurate. Bias in responses also complicates the interpretation of time trends, as changes in measured LGBT-related sentiment could be true changes in sentiment or changes in reporting. Improving the methods of asking such questions is valuable not only for the measurement of LGBT-related sentiment, but potentially also for gathering data on sentiment regarding policies related to race and gender. The economics literature on discrimination has typically avoided asking about beliefs directly, instead relying on the observation of behavior. 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). Augmenting or linking the existing evidence on behaviors to accurate direct evidence on beliefs could prove quite useful.

Individuals are reluctant to respond honestly on surveys in a variety of contexts because they want their answer to adhere to social norms—a phenomenon known as “social desirability bias”

---

<sup>1</sup> For a review, see Black, Sanders, and Taylor (2007).

<sup>2</sup> 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.

<sup>3</sup> For instance, Black, Gates, Sanders, and Taylor (2002) examine the location decisions of the LGBT population.

<sup>4</sup> See Bloom and Glied (1992), see Berg and Lien (2006), Black, Gates, Sanders and Taylor (2000), and Fay, Turner, Klassen and Gagnon (1989).

<sup>5</sup> 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.

<sup>6</sup> See Li and Nagar (2013) on the LGBT diversity in the workplace, and Klawitter and Flatt (1998) on the effect of antidiscrimination policies.

(Maccoby and Maccoby 1954, Edwards 1957, Fisher 1993). 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.

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 1998). The use of computers as intermediaries between interviewers and survey participants, as well as self-administered questionnaires have proven to be an important step forward in increasing privacy and are considered current best practices (Turner et al. 1998, Williams Institute 2009). 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).<sup>7</sup> However, it is unknown how accurate these data are.

We run an experiment to test whether anonymity and privacy are sufficient for eliciting truthful responses to questions about sexuality. Put differently, we ask whether current best practices eliminate social desirability bias and elicit truthful reporting. We find substantial underreporting of LGBT identity and behaviors as well as underreporting of anti-gay sentiment intolerance even under anonymous and very private conditions. Finally, we show that when data are collected in a way that prevents inference about any particular individual, a concept we will refer to as *individual inference*, the quality of our estimates is significantly improved.

We adopt a method proven to reduce social desirability bias, the item count technique (ICT) (Miller 1984). (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, where the items are neutral and non-sensitive in nature. The rest of the respondents report how many of  $N+1$  items are true, with  $N$  items being identical to the control group's items, and the  $N+1$ st item being a sensitive item, e.g. “I am not heterosexual”.<sup>8</sup> With a large enough sample, the researcher can estimate the population mean for the  $N+1$ st item, the sensitive item, 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 individual inference. A commonly-used privacy criterion for information stored in databases is “differential privacy”— a guarantee that information released from a database cannot uniquely identify an individual (Dwork 2008, Wasserman and Zhou 2010.

---

<sup>7</sup> Outside the U.S. context, recent biometric validation in Zimbabwe of computer-assisted self-interviews showed underreporting for sensitive sexual questions (Minnis et al. 2009).

<sup>8</sup> “ $N+1$ st” does not indicate the sensitive item will necessarily be the last item shown in the list.

See Heffetz and Ligett 2013 on applications to economics). Often, to implement this guarantee, information from the database is released with noise added. The ICT method goes one step further in protecting privacy: by eliciting the information in a coarser way, the database does not ever contain any particular individual's answer to the sensitive question.

The ICT has typically been used to reduce the psychological cost of admitting an unacceptable behavior to an interviewer: Saying “three items” might be easier to say than “Yes, I cheat on my spouse.” Our experiment does not have an interviewer in any treatment arm. Rather, our design uses the ICT to eliminate precise inference about any given observation in our dataset in an already private and anonymous setting. Note that the data collected is anonymous in any case, so that even apart from the ICT, the researcher could only make inferences about a particular record in the dataset (e.g. observation #20), not about an individual. The ICT goes a step further by removing the possibility of identifying sensitive facts about any given observation (i.e. the researcher cannot even tell if observation #20 engaged in the sensitive behavior).

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  $N$  non-sensitive 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% ( $p < 0.05$ ), same-sex sexual experiences by 59% ( $p < 0.01$ ), and directionally same-sex attraction by 9.4% (n.s.). We combine all own-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.<sup>9</sup> Respondents were 67% more likely to express disapproval of an openly gay manager at work ( $p < 0.01$ ), 71% more likely to say it should be legal to discriminate in hiring on the basis of sexual orientation ( $p < 0.01$ ), 22% less likely to support the legality of same-sex marriage, 46% less likely to support adoption by same-sex couples ( $p < 0.10$ ), and 32% *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 ( $p < 0.01$ ). Taken together, these results indicate that both non-heterosexuality and anti-gay sentiment are substantially underestimated in existing surveys.

---

<sup>9</sup> There may also be implicit discrimination (Bertrand & Mullainathan 2005), which the ICT does not necessarily capture.

Our comparison between Veiled and Direct Report allows us to identify social norms: as the veil is eliminated, responses move towards what is more socially acceptable. Our results are 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 find evidence that many individuals remain uncomfortable reporting non-heterosexual identity and behavior.

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. While our sample is non-representative of the U.S. population, it is important to note that we under-sample groups for whom we generally estimate relatively larger treatment effects. This suggests that the extent of underreporting may be even higher for the general population than we report here.

## Existing Literature

### *Existing Measures of the LGBT Population*

Research on the LGB<sup>10</sup> population in the U.S. has been hindered due to lack of data availability, as few representative surveys ask about sexual orientation. The literature faces two challenges: measuring the size of the LGB population and its characteristics. Ultimately, researchers would like to analyze the characteristics of the LGB population (i.e. household formation, human capital investment, geographic location, and labor market outcomes). However, first—and most relevantly—researchers must identify who in the population is LGB. In doing so, they produce “incidence rates” (i.e. the fraction of the population that is LGB). Surveys vary in many factors that can affect measured 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. Furthermore, incidence rates vary depending on the definition used to classify individuals: e.g. based on sexual desire, sexual practices (over various time horizons), or self-announced identity.

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.<sup>11</sup> For self-identification as LGB, estimates range from 1.7% of adults (National Epidemiological Survey on Alcohol and Related Conditions, 2004-2005) to 5.7% of adults (National Survey of Sexual Health and Behavior 2009). Identification as gay or lesbian is relatively more stable across surveys (ranging from 1% to 2.5%) than identification as bisexual (0.7% to 3.1%).

---

<sup>10</sup> 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% of the population identifies as LGB and only 0.3% as transgender.

<sup>11</sup> The early work of Kinsey, Pomeroy and Martin (1948) was not based on a representative sample of the population.

Other ways of measuring sexual orientation produce much higher rates. The National Survey of Family Growth (NSFG) provides a good illustration of the issues raised in measuring the LGB population. Conducted by the CDC's National Center for Health Statistics, it interviews 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% of adults reported any same-sex attraction, and 8.8% of adults reported any same-sex sexual behavior; the fraction of adults identifying as LGB was 3.7% in that survey.

Moreover, existing measures offer a range of gradations. The NSFG allows participants to classify sexual attraction from attraction only to opposite sex, mostly to opposite sex, equally to both, mostly to same-sex, only to same-sex, or not sure. Substantial gender differences are found: 13.3% of women versus 7.8% of men classified themselves in one of the categories other than "only opposite sex" (only a small fraction, 0.7-0.8% said not sure). We replicate these gender differences in our Direct Report treatment.

Data on the characteristics of the LGBT population, outside of sexual behavior, is quite limited. Black, Sanders, and Taylor (2007) give an overview of the available data on the LGB population's characteristics other than sexual behavior. While the Census is a natural source of data for many demographic characteristics, it cannot be used to give estimates of the size of the LGBT population because it does not ask about sexual orientation. However, beginning in 1990, the Census can be used to identify the subset of LGBT individuals who are partnered with someone of the same sex ("unmarried partner") and who are willing to disclose this fact to the Census. When interpreting the Census (and the later American Community Survey), researchers need to be cognizant of issues raised by political controversy and measurement error.<sup>12</sup>

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% say they have been discriminated against by an employer and 30% say they have been physically attacked or threatened.

### *Validation of the ICT*

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), and it generally increases respondents'

---

<sup>12</sup> For the census data, measurement error must be addressed (Black, Gates, Sanders, and Taylor. 2007), and the data itself was systematically recoded for individuals who indicated they were in a same-sex marriage during certain time periods. The Census saw protests from LGBT activist groups against not being counted ("Queering the Census"), as well as disagreement over the interpretation of "partnership" and "marriage": i.e. same-sex marriage was legal in certain states, but not recognized by the federal government, and individuals might be in a committed relationship but not consider it a "partnership". See Virgile (2011) for more detail on measurement and history.

perception of privacy, as compared to other computer-aided elicitation (Coutts and Jann 2011). 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 and Karlan 2012). The ICT has never been applied to measure sexual orientation of the general population, and rarely to measure opinions on public policy.

Importantly, 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).

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.

## **Experiment Design**

We investigate eight questions, detailed in Table I. 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. 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.



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).

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" and "Veiled Report" (a between-subject design). 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 innocuous items, common across treatments.

For two reasons, each set of four innocuous 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. Second, the negative correlations also decrease the odds 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.

In the Direct Report condition, participants first saw a list of four innocuous statements and were asked to indicate how many of the four statements were true for them. 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 innocuous 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 I, Panel B or in the reverse ordering. In Table A3, 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). This may be because participants perceived Question 8 as less sensitive than Question 1.

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 A4 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.

Data was collected in two waves. The first wave was conducted from November 1<sup>st</sup> – November 3<sup>rd</sup>, 2012, just prior to the United States’ presidential election; 786 individuals participated over this three-day window. The second wave was conducted just after the presidential election, from November 7<sup>th</sup> – November 15<sup>th</sup>, 2012; 1730 individuals participated during this window.<sup>13</sup>

The design of the Direct Report treatment was slightly different in the first wave. Within each question, one innocuous statement was separated from the list and asked directly. This difference is illustrated in Table A5 in the Appendix. 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 innocuous 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, in fact, affect the results presented in the main text: in Table A6 in the Appendix, 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.

## 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 innocuous 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.

---

<sup>13</sup> 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. An original goal was to identify 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, and we ultimately failed to have enough power to make meaningful comparisons.

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 as  $\mu \equiv E[y_{qi}^V] - E[y_{qi}^D]$ . We can also interpret  $\mu$  as a measure of how stigmatized the sensitive response is; a larger  $\mu$  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  $\mu$ , as it allows us to control for observed demographics. Thus, in our results below, we will report the estimated  $\mu$  from the regression:

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

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).

## Experiment Results

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 II provides descriptive statistics. Our sample is approximately 42% female with a median age of 26. Less than 32% describe themselves as being at least moderately religious, and less than 16% self-reports as Republican. More detailed descriptive statistics, including descriptive statistics broken out by treatment, are available in Appendix Table A1. Attrition in the experiment was very low (2.97% of participants assigned to treatment), and did not differ significantly by treatment.<sup>14</sup> 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.

---

<sup>14</sup> 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 III would be reduced by only 1 percentage point.

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 would 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 do not substantially change (see Appendix).

Before turning to our regression results, we first present the histograms of responses to each of our questions. In Figure 1 below, 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% 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. Second, the distributions of  $y_{qi}$  are very clearly non-uniform. More importantly, if we compare the distributions across treatment 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.

Table III 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,  $\mu$ , as a percent of the total sample, estimated using a regression with controls described in the empirical approach section. (Note that  $\mu$  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.<sup>15</sup>

We present heteroskedasticity-robust standard errors for the treatment effect  $\mu$  in Table III. When calculating the percent increase in respondents answering yes to the sensitive question (Table III, column 4) and the estimated true fraction answering yes to the sensitive question (Table III, column 3), we use the bootstrap to calculate standard errors (estimated 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  $\mu$  are quite similar to the heteroskedasticity-robust ones reported.

### *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

---

<sup>15</sup> The treatment effects estimated by our regression specification are very similar to the treatment effects that would be estimated by a simple comparison of means across treatment (see Appendix Table A2).

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. In the Direct Report treatment, 11% of the population reports that they do not consider themselves heterosexual (8% for men, 16% for women). In the Veiled Report treatment, this increases to 19% (15% for men, 22% for women). The 7.3 percentage point difference is significant at  $p < 0.05$ , and represents a 65% increase in the fraction of the sample reporting as non-heterosexual.

In “Question 3-Experience”, the number of participants reporting having had a sexual experience with someone of the same sex increases from 17% (12% for men, 24% for women) in the Direct Report treatment to 27% (17% for men, 43% for women) in the Veiled Report treatment, a 59% 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. See Appendix for more details.<sup>16</sup>)

Finally, we create an “own sexuality index” for each individual by summing the answers to each of the separate own sexuality questions. We code the questions so that positive answers indicate sensitive answers, as described in Table I. 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 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 IV 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 ( $p < 0.01$ ).

---

<sup>16</sup>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.

### *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% of the Direct Report treatment did *not* support the legal recognition of same-sex marriages. This increases to 23% 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% in the Veiled Report treatment compared to the Direct Report treatment, from 16% to 27% ( $p < 0.01$ ). “Question 6-Discriminate”, asks whether the respondent believes it should be illegal to discriminate in hiring based upon sexual orientation. While only 14% in the Direct Report treatment say that this type of discrimination should *not* be illegal, in the Veiled Report treatment, we estimate that 25% 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% opposed) than in the Direct Report treatment (13% 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% under Direct Report to 15% ( $p < 0.05$ ). This indicates that participants saw it more socially desirable to report that sexual orientation is changeable, which goes in the opposite direction of a general “pro-LGBT” norm.

Just as for the “own sexuality” questions, we sum the answers of the 5 sentiment questions to create an overall sentiment index (see Table IV). Here, the questions are coded so that positive answers indicate sensitive answers (anti-LGBT sentiment). In this index, the number of sensitive anti-gay sentiment answers<sup>17</sup> rises by 0.24 in Veiled Report condition (two-sample Wilcoxon rank-sum test,  $p < 0.01$ ).

### *Treatment Response by Demographics*

Table V 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

---

<sup>17</sup> 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.

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% to 21%) in Question 1 ( $p < 0.05$ ) and same-sex sexual experiences by 14 percentage points (from 11% to 25%) in Question 3 ( $p < 0.05$ ), compared to the Direct Report. These are increases of 163% and 127%, 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% to 30% ( $p < 0.01$ ), a 233% increase, and the fraction reporting a same-sex experience increases from 18% to 38% ( $p < 0.05$ ), a 111% 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 VI, we present the treatment effects for Questions 4-8 by demographic groups. Again, Direct Report responses can be found in the Appendix 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 conformation 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% in the Direct Report, 54% using the Veiled Report), an insignificant difference. For Democrats, the treatment effect is larger, with our estimate of non-supporting Democrats increasing from 10% to 20% 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% to 67% ( $p < 0.01$ ). When asked directly, only 23% 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% ( $p < 0.01$ ). 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.

## Discussion

In sum, estimates using a Veiled Report elicitation - one that precludes inferences at the individual level - show that standard methods of eliciting respondents' sexual orientation and behavior underestimate the true fraction of individuals who do not identify as heterosexual and who have had a same-sex sexual encounter. Our population has broad coverage of demographic characteristics, but is not representative of the U.S. as a whole (e.g., 18-30 year-old liberals are overrepresented in our sample). Thus, while our results do indicate that existing surveys substantially underestimate the size of the LGBT population and magnitude of anti-gay sentiment, our results should not be interpreted as giving the true fraction of non-heterosexuality in the U.S. Nonetheless, the correlates and level of misrepresentation presented here may be useful for other researchers to estimate the extent of bias in their data.

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.

The misreporting of sexual identity and sexuality-related opinions that we observe has far-reaching implications. 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?). If individuals' LGBT identity is underreported, it suggests that other items related to that identity would also be underreported: for instance, data on workplace or housing discrimination or hate crimes. 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, the cost-benefit analysis of LGBT-related public health interventions, elder services, workplace policies, domestic violence prevention programs, and youth mental health/suicide prevention programs.

Our finding that there is stigma attached to reporting anti-gay sentiments is perhaps even more 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.<sup>18</sup> The fact

---

<sup>18</sup> 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 these opinions are still misrepresented suggests that many other opinions on controversial public issues may not be accurately measured.

## 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.

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>

Cochran, Susan D., J. Greer Sullivan, and Vickie M. Mays. "Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay, and bisexual adults in the United States." *Journal of consulting and clinical psychology* 71.1 (2003): 53-61.

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.

Dwork, Cynthia. "Differential privacy: A survey of results." *Theory and Applications of Models of Computation*. Springer Berlin Heidelberg, 2008. 1-19.

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).

Gavrilets, Sergey, and William R. Rice. "Genetic models of homosexuality: generating testable predictions." *Proceedings of the Royal Society B: Biological Sciences* 273.1605 (2006): 3031-3038.

Heffetz, Ori and Katrina Ligett. 2013. "Privacy and data-based research." *NBER Working Paper* #19433.

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, and Dean Karlan. "Measuring Preferences and Predicting Outcomes." *Working Paper Yale University*, (2011).

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.

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.

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.

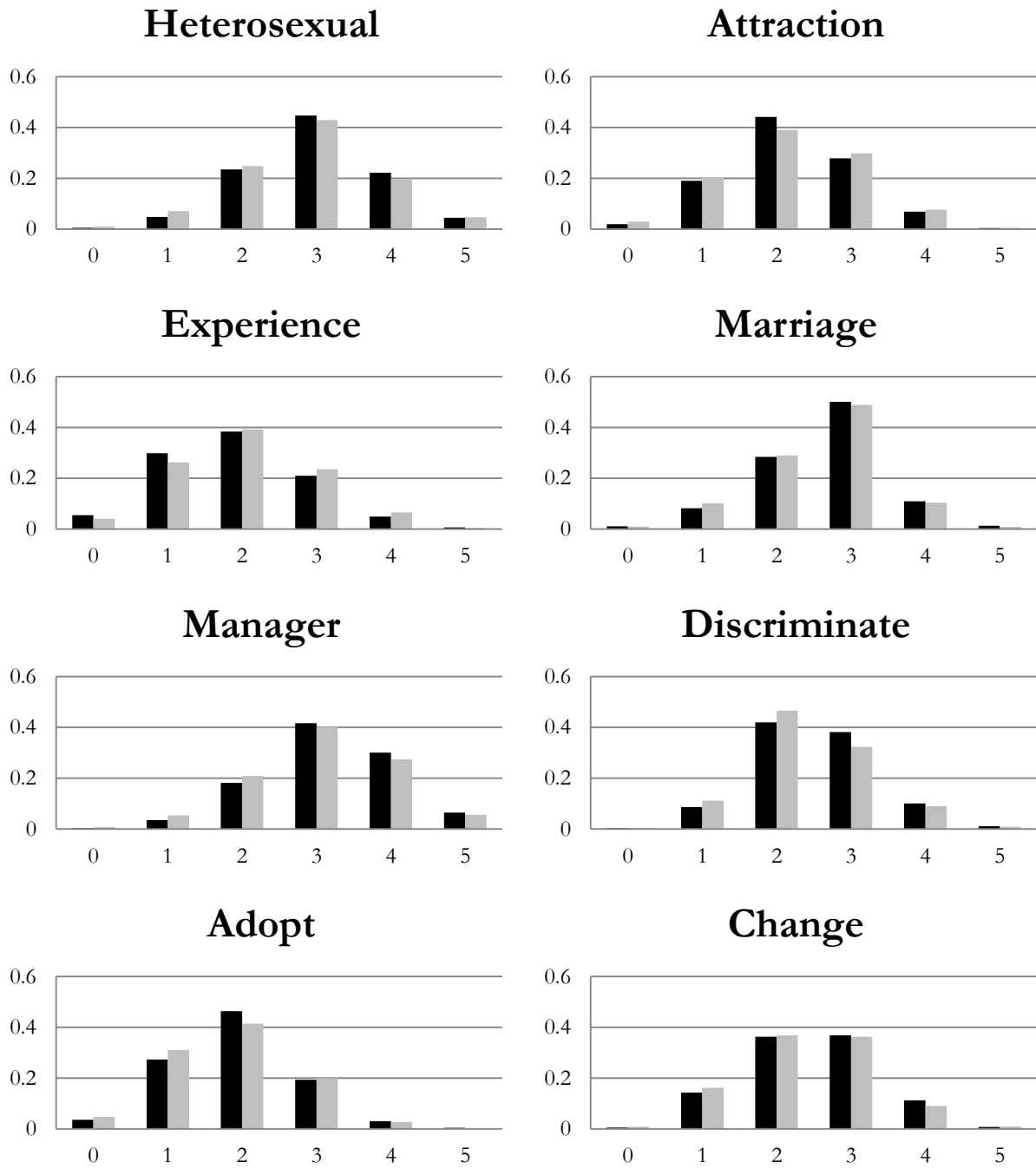
- Lundberg, Shelly, and Robert A. Pollak. 2007. "The American Family and Family Economics" *Journal of Economic Perspectives*, 21(2): 3-26.
- Maccoby, E. E. and N. Maccoby. "The Interview: A tool of social science." In G. Lindzey (Ed.), *Handbook of Social Psychology: Vol. 1 Theory and Method*. Cambridge, 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)
- Minnis, Alexandra M., et al. "Biomarker validation of reports of recent sexual activity: results of a randomized controlled study in Zimbabwe." *American Journal of Epidemiology* 170.7 (2009): 918-924.
- 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/>
- 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.
- 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.
- Turner, Charles F., et al. "Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology." *Science* 280.5365 (1998): 867-873.
- Virgile, Matt. "Measurement Error in the Relationship Status of Same-Sex Couples in the 2009 American Community Survey." 2011 American Community Survey Research and Evaluation Report Memorandum Series ACS11-RER-10 (2011).

Wasserman, Larry, and Shuheng Zhou. "A statistical framework for differential privacy." *Journal of the American Statistical Association* 105.489 (2010): 375-389.

Williams Institute, "Best Practices for Asking Questions about Sexual Orientation on Surveys" (2009)

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.

# Tables and Figures



**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 I: Experimental Design**

**Panel A: Comparison of Direct Report and Veiled Report Treatments**

Direct Report	Veiled Report
<ul style="list-style-type: none"> <li>● I remember where I was the day of the Challenger space shuttle disaster.</li> <li>● I spent a lot of time playing video games as a kid.</li> <li>● I would vote to legalize marijuana if there was a ballot question in my state.</li> <li>● I have voted for a political candidate who is pro-life.</li> </ul> <p>Please fill in the bubble that corresponds to the total number of statements above that apply to you.</p> <p style="text-align: center;">0      1      2      3      4</p> <p>Do you consider yourself to be heterosexual?</p> <p style="text-align: center;">Yes    No</p>	<ul style="list-style-type: none"> <li>● I remember where I was the day of the Challenger space shuttle disaster.</li> <li>● I spent a lot of time playing video games as a kid.</li> <li>● I would vote to legalize marijuana if there was a ballot question in my state.</li> <li>● I have voted for a political candidate who is pro-life.</li> <li>● I consider myself to be heterosexual.</li> </ul> <p>Please fill in the bubble that corresponds to the total number of statements above that apply to you.</p> <p style="text-align: center;">0      1      2      3      4      5</p>

**Panel B: Sensitive Questions Used**

Question	Sensitive Answer
<i>Own Sexuality</i>	
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”
<i>LGBT-related Sentiment</i>	
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”

**Table II. Descriptive Statistics**

Age: 30 and under	66%
Age: 31-50	28%
Age: over 50	6%
Gender= Male	57%
White	80%
Black	6%
Other	9%
Married	28%
Is Parent	29%
Works Full Time	38%
Works Part Time	18%
Current Student	22%
Northeast Census Region	20%
Midwest Census Region	22%
South Census Region	34%
West Census Region	24%
<i>Education Category</i>	
Finished High School	11%
Some College	43%
Finished College	31%
Graduate School (some or more)	14%
<i>Religion</i>	
Christian	36%
Jewish	2%
No religion	43%
Mean response to "how religious are you" on 1-7 scale	2.67 (2.00 SD)
<i>Political Views</i>	
Republican	16%
Democrat	46%
Independent	32%
Mean response to "how political are you" on 1-7 scale	4.37 (1.73 SD)



**Table III. The Effect of Veiled Report Treatment on Reports of Sensitive Behaviors**

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

Notes: \*n= 2516, with 1270 in Direct Ask condition. Column 1 is the sample mean. Column 2 is the coefficient  $\mu$  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.

**Table IV. 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 <sup>2</sup>	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  $\mu$  on “Veiled Report” from a regression with controls. Heteroskedasticity-robust standard errors in brackets.

**Table V. 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
<i>Gender</i>					<i>Politics</i>				
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]	
<i>Race</i>					<i>Age</i>				
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]	
<i>Religion</i>					51 plus				
Christian	12.9	4.51	13.8	905		9.22	3.79	32.3	158
	[6.46]	[6.03]	[6.25]			[15.7]	[18.7]	[17.1]	
No Religion	0.0204	5.01	3.71	1078					
	[4.87]	[5.55]	[5.85]						

Note: n= 2516. The table gives the coefficient  $\mu$  on “Veiled Report” from a regression with controls run on each demographic subgroup, equivalent to Column 2 of Table 1. (Details in Appendix). Heteroskedasticity-robust standard errors in brackets.

**Table VI. Change in Sensitive Answer Reports for Opinions Questions in Veiled Report, by Demographics**

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

Note: n= 2516. The table gives the coefficient  $\mu$  on “Veiled Report” from a regression with controls run on each demographic subgroup. (Equivalent to Column 2 of Table 1.) Heteroskedasticity-robust standard errors in brackets.

# Appendix

**Table A1. Detailed Descriptive Statistics**

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

Midwest	22%	23%	22%
South	33%	35%	34%
West	25%	24%	24%
<i>Education</i>			
Some High School	1%	2%	1%
Finished High School	11%	11%	11%
Some College	44%	43%	43%
Finished College	30%	32%	31%
Some Graduate School	5%	5%	5%
Finished Graduate School	9%	8%	9%
<i>Religion</i>			
Christian	37%	36%	36%
Jewish	2%	2%	2%
No Religion	42%	44%	43%
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)
<i>Political Views</i>			
Republican	15%	17%	16%
Democrat	46%	46%	46%
Independent	33%	31%	32%
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)
<i>Answered the Cognitive Reflection Task Correctly</i>	39%	40%	40%

---

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

**Table A2. Comparisons of Means Across Treatment**

	Mean Response in Direct Report (SD)	Mean Response in Veiled Report (SD)
Question 1 - Heterosexual	2.97 (0.930)	2.87 (0.992)
Question 2 - Attraction	2.20 (0.902)	2.20 (0.962)
Question 3 - Experience	1.92 (0.985)	2.04 (0.978)
Question 4 - Marriage	2.65 (0.867)	2.60 (0.870)
Question 5 - Manager	3.17 (0.937)	3.05 (0.990)
Question 6 - Discriminate	2.52 (0.838)	2.41 (0.843)
Question 7 - Adopt	1.92 (0.872)	1.86 (0.894)
Question 8 - Change	2.46 (0.915)	2.39 (0.921)

**Table A3. Order Effects**

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



**Table A4. 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
<i>Own Sexuality</i>				
Not Heterosexual	11.3 [0.89]	10.8 [1.03]	7.3 [3.57]	7.7 [4.14]
Same sex Attraction	13.9 [0.97]	13.8 [1.14]	1.3 [3.63]	2.8 [4.21]
Same-sex sexual experience	17.2 [1.06]	16.2 [1.22]	10.1 [3.82]	12.4 [4.41]
<i>Attitudes and Opinions about Issues Related to Sexual Orientation</i>				
Not Support Same-sex Marriage	18.8 [1.10]	19.5 [1.31]	4.2 [3.18]	0.7 [3.85]
Not Happy with LGB Manager	16.2 [1.03]	17.0 [1.24]	10.8 [3.75]	16.4 [4.62]
Not Illegal to Discriminate	14.4 [0.99]	13.2 [1.12]	10.3 [3.33]	11.5 [4.08]
LGB not allowed to Adopt	12.9 [0.94]	13.1 [1.12]	5.9 [3.40]	7.6 [3.80]
Can Change Orientation	22.2 [1.17]	24.0 [1.41]	-7.0 [3.52]	-5.8 [4.33]

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

**Table A5. Direct Report Design Differences Across Wave**

First Wave Direct Report	Second Wave Direct Report
<ul style="list-style-type: none"> <li>● I remember where I was the day of the Challenger space shuttle disaster.</li> <li>● I spent a lot of time playing video games as a kid.</li> <li>● I would vote to legalize marijuana if there was a ballot question in my state.</li> </ul> <p><b>Please fill in the bubble that corresponds to the total number of statements above that apply to you.</b></p> <p style="text-align: center;">0      1      2      3</p> <p><b>Have you voted for a political candidate who is pro-life?</b></p> <p style="text-align: center;">Yes   No</p> <p><b>Do you consider yourself to be heterosexual?</b></p> <p style="text-align: center;">Yes   No</p>	<ul style="list-style-type: none"> <li>● I remember where I was the day of the Challenger space shuttle disaster.</li> <li>● I spent a lot of time playing video games as a kid.</li> <li>● I would vote to legalize marijuana if there was a ballot question in my state.</li> <li>● I have voted for a political candidate who is pro-life.</li> </ul> <p><b>Please fill in the bubble that corresponds to the total number of statements above that apply to you.</b></p> <p style="text-align: center;">0      1      2      3      4</p> <p><b>Do you consider yourself to be heterosexual?</b></p> <p style="text-align: center;">Yes   No</p>

**Table A6. Estimated Treatment Effects by Wave of Data Collection**

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

**Table A7. Direct Ask Estimates for Own Sexuality Questions, by Demographics**

	1-Hetero.	2-Attract.	3-Experience	N		1-Hetero.	2-Attract.	3-Experience	N
<i>Gender</i>					<i>Politics</i>				
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]	
<i>Race</i>					<i>Age</i>				
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]	
<i>Religion</i>					51 plus				
Christian	0.0821	0.0994	0.112	463		0.0506	0.0380	0.127	79
	[0.0132]	[0.0137]	[0.0152]			[0.0246]	[0.0190]	[0.0396]	
No Religion	0.135	0.164	0.230	535					
	[0.0152]	[0.0164]	[0.0186]						

**Table A8. Direct Ask Estimates for Sentiment Questions, by Demographics**

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

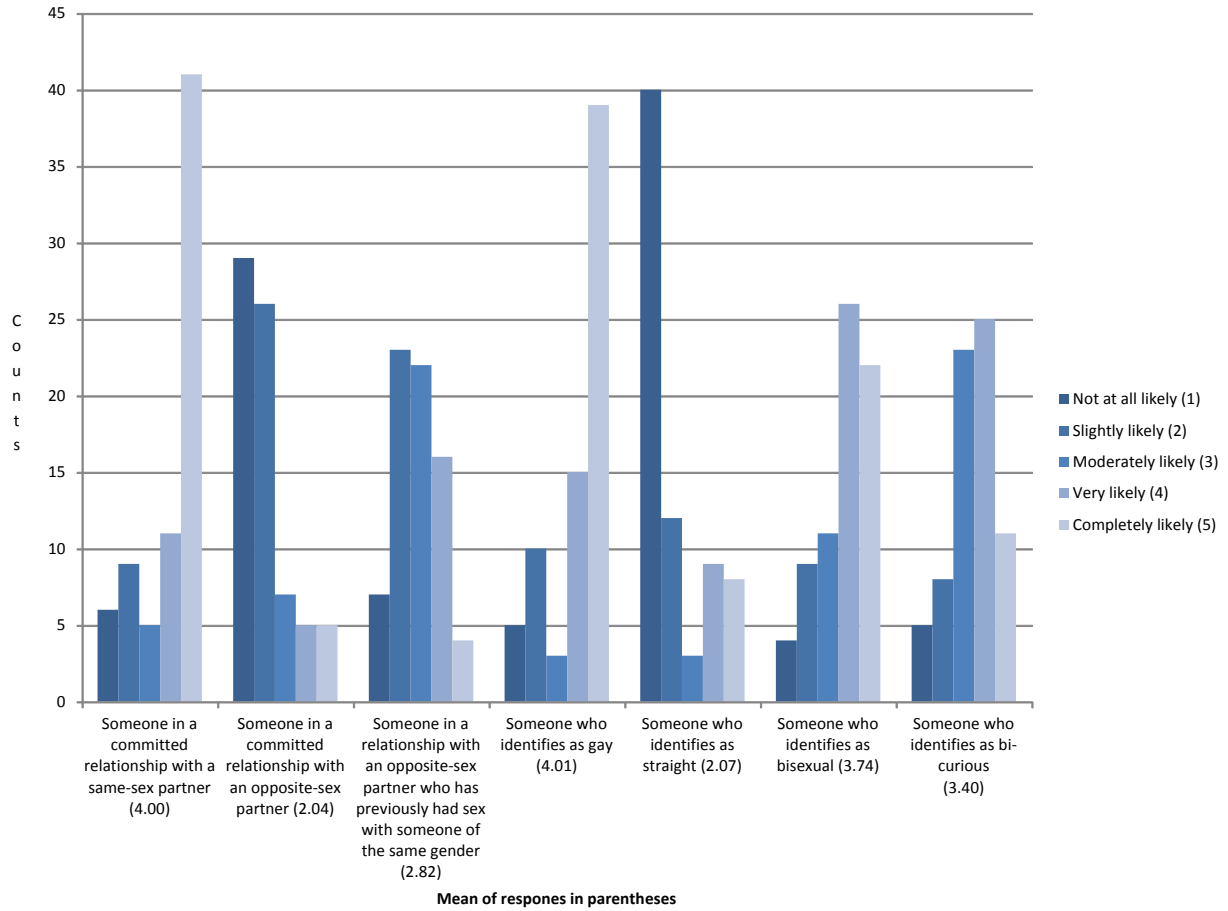


Figure A1. *How likely is it that the following people would answer YES to the question, "Are you sexually attracted to members of the same sex?"*