PRE-ANALYSIS PLAN: DO "SAFE SPACES" REDUCE SENSITIVITY BIAS?

Lisa Mueller*

April 17, 2019

Contents

*Assistant Professor, Department of Political Science, Macalester College. lmueller@macalester.edu.

1 Motivation

Questionnaire-based research, encompassing conventional surveys as well as many experiments, occurs in a social context. It is interpersonal even when enumerators do not interview subjects face-to-face, because respondents presumably understand that another human wrote the questions and will read their answers (Berinski 2004).[1] Researchers should therefore consider how social norms may bias response patterns. A specific concern is *social desirability bias*, or "the tendency on behalf of subjects to deny socially undesirable traits and to claim socially desirable ones, and the tendency to say things which place the speaker in a favourable light" (Nederhof 1985, 264). Political scientists have recently turned their attention to status motivations—that is, how people seek to distinguish themselves from others along dimensions of income, assets, attributes, actions, or achievement (McClendon 2018). However, people sometimes do the opposite, trying to *conform* with majority opinions and behaviors in order to avoid social awkwardness (in mild cases) or ridicule and physical reprisal (in more extreme cases) (Cialdini and Trost 1998). This "impression management" is of intrinsic interest to psychologists (Krumpal 2013), but also has ramifications for the quality of data on any number of topics; data gathered with questionnaires can impede valid inference if they reflect the thoughts of groups rather than those of individual respondents.[2] Moreover, policymaking is difficult without accurate data on rates of phenomena that people hesitate to discuss openly, including abortion, assault, and attitudes toward existing controversial policies. If people from nondominant groups are especially prone to under-report such phenomena, as previous findings suggest (Odendaal 2015; Johnson and Van de Vijver 2015; Bernardi 2006), then

---

[1]The present study will verify this with a question at the end of a survey asking respondents who they think commissioned the study. See the complete survey instrument in Section 8.

[2]Social desirability bias thus relates to *groupthink*, "a mode of thinking that persons engage in when *concurrence-seeking* becomes so dominant in a cohesive ingroup that it tends to override realistic appraisal of alternative courses of action" (Janis 1971, 84). Groupthink is different, though, because it refers to deliberative decision-making. In contrast, survey respondents generally choose answers alone.

policies may under-serve communities that most need assistance.[3]

Social desirability bias is often conflated with several other forms of bias, as Blair, Coppock and Moor (2018, 7) explain. First, pressure to answer a question in a particular way could stem from respondents themselves, rather than from observers. Second, respondents might consider questions too intrusive or taboo to answer at all, regardless of their concern for the social desirability of specific answers. Third, respondents might give dishonest answers out of fear that authorities or rivals will hear them. Finally, respondents might feel pressure to represent their communities, as opposed to themselves, in a particular light—what is known as "sociotropic misreporting." Blair, Coppock and Moor (2018) propose the term *sensitivity bias* to synthesize these related forms of bias. It is the term I use in this study.

Sensitivity bias has received substantial attention in the psychology and public health literatures (Hebert et al. 1997; Krumpal 2013) but less in political science,[4] where survey questions can be just as sensitive. Political scientists may be less likely to pose intrusive questions about diet, drug use, or sexual habits, but they frequently inquire about candidate preference, participation in violence or corruption, and attitudes on social issues like gay rights, which could expose truthful respondents to social sanctioning or personal shame.

Social sensitivity is a *subjective norm*, or what a person believes she should say or do to please a specific referent. That referent could be a survey enumerator, or—if the survey occurs in public—bystanders. Subjective norms can clash with *personal norms*, meaning one's own internalized standards for behavior (Cialdini and Trost 1998). For example, a respondent could harbor a subjective norm of appearing politically centrist but a personal norm of "speaking truth to power" by calling out perceived social injustices. Which norm will ultimately have more sway over how the respondent answers a politically sensitive question? Existing research suggests that it depends in

---

[3]In some cases, people from dominant, rather than nondominant, groups are more prone to socially desirable responding. For example, men are more reluctant than women to report domestic abuse (Felson and Paré 2005).

[4]See a meta-analysis by Blair, Coppock and Moor (2018) for key exceptions.

part on the design and setting of the survey, which can prime the respondent to worry about how others will judge his response, thus eliciting a vague or socially desirable answer instead of a sincere one. In other words, "When norms compete, situational signals might activate one norm over the other" (Cialdini and Trost 1998, 161).

Adida et al. (2016) studied sensitivity bias in surveys of African societies, where subjective norms often dictate that people not show ethnic antipathies or favoritism. These norms are particularly strong if the enumerator and respondent are of different ethnic identities, heightening fear of judgment. In contrast, when an enumerator interviews a coethnic, the respondent is more comfortable admitting that she feels strongly attached to her ethnic group, that her group is economically and political disadvantaged, and that she disapproves of a president from another ethnic group. Similar patterns emerge in mixed settings. In the United States, where race is a politically salient identity, white people express more support for interracial marriage when surveyed by black rather than white enumerators (Hatchett and Schuman 1975). Similarly, Americans express more pro-labor attitudes when the interviewer is working-class (Katz 1942). In Egypt, Muslims present as more religious on surveys administered by someone wearing an Islamic headscarf (Blaydes and Gillum 2013).

Scholars are wise to warn against the pitfalls of "interviewer effects," but this is only one potential source of sensitivity bias. Others include question wording and the timing of survey enumeration. A sketch on the comedic television show *Jimmy Kimmel Live* made fun of Americans who oppose "Obamacare" but praise "the Affordable Care Act" (both are names for the same policy). The ability of question wording to bias answers underlies the method of survey experiments, where researchers randomize versions of a question to ascertain causal effects on respondent opinion (e.g., Hainmueller and Hiscox 2010). Studying how survey timing influences sensitivity bias, Karp and Brockington (2005) found that people in five countries are more likely to over-report voting during elections that draw high public attention and voter turnout, because the pressure to vote in such settings is greater than in low-profile elections. Adida et al. (2018) likewise found

4

more over-reporting of turnout and voting for the incumbent in more competitive and rural areas of Benin.

Scholars have applied various techniques to address sensitivity bias. These include wording questions as neutrally as possible, randomizing questions or response options, reassuring respondents of their anonymity, asking validation questions (the answers to which only honest respondents would know), leading respondents to believe that a device will detect whether they are lying, matching the identities of enumerator and respondent, and having respondents administer questionnaires to themselves (thus avoiding face-to-face interaction) (Nederhof 1985; Blair, Imai and Zhou 2015; Rosenfeld, Imai and Shapiro 2016). A popular tool in political science is the list experiment, or unmatched item count method, where enumerators present a subject with one of two lists of response options, one of which contains a socially sensitive answer (e.g., "I fought in the rebellion."). Respondents report only the number of response items that apply to them, not their specific answers. Because the list that a respondent receives is assigned at random, any difference in the average number of response items across treatment and control groups can be attributed to differences in the proportion of respondents who exhibit the socially sensitive characteristic. Political scientists have used list experiments to study vote buying (Gonzalez-Ocantos et al. 2012), racism (Kuklinski et al. 1997), voter turnout (Holbrook and Krosnick 2009), and other variables. While list experiments are more elegant than many other ways of avoiding sensitivity bias, there is no guarantee that including a sensitive item in a list will not change responses to control items (constituting *design effects*), or that respondents will not lie about the sensitive item despite confidentiality protections (Blair and Imai 2012, 62). List experiments are also highly imprecise; direct questions, biased though they may be, are often preferable unless sensitivity bias is extreme (Blair, Coppock and Moor 2018, 3). Alternative techniques address sensitivity bias post-hoc, rather than trying to prevent it. For example, scholars have used scales for measuring sensitivity bias (see Section 3) and then omitted data from high-scoring subjects, or simply mentioned the presence of bias when reporting results (Nederhof 1985, 268). These solutions are imperfect, leaving the

door open to exploring new ones. The present study investigates whether it is possible to mitigate sensitivity bias through a heretofore untested intervention: creating "safe spaces" for respondents.

A safe space is "an environment in which everyone feels comfortable expressing themselves and participating fully, without fear of attack, ridicule, or denial of experience" (Arao and Clemens 2013, 138). The concept arises usually in Western classroom settings where teachers try to engage students in frank, respectful discussions about sensitive issues. Creating safe spaces, also called "positive spaces" in Canada, involves setting ground rules such as, "Students are able to openly express their individuality, even if it differs dramatically from the norms set by the instructor, the professor, or other students" (Holley and Steiner 2005, 50). By establishing a similar ground rule at the start of a survey, above and beyond typical reassurances of anonymity (see Section 2.1), researchers might be able to make respondents overcome their proclivity toward sensitivity bias. Showing the effectiveness of safe-space language would not nullify existing tools for reducing sensitivity bias, but rather provide an additional and less technical option for researchers. Other tools could still be most appropriate in certain contexts, like in war zones where protecting respondents from retaliation requires completely obscuring individual responses to sensitive questions by using a list experiment or endorsement experiment (e.g., Blair, Imai and Lyall 2014). On the proverbial minefield of contemporary Western college campuses, the classroom is still a relatively safe space even without instructors tailoring their language.

The present study comprises a novel application of safe-space language to questionnaire-based research, building on applications to qualitative methods including participatory action research (Stoudt 2007) and focus groups (Kamberelis and Dimitriadis 2011). It furthermore contributes objective, empirical insights to politically charged debates about whether safe spaces promote or stifle free speech on college campuses; responding honestly to a questionnaire is one form of speaking freely. Some people argue that safe spaces encourage free expression by allowing students from non-dominant groups to participate in class without fear of emotional, social, or physical harm (Downes 2016). For example, the Office of LGBTQ Student Life at the University of

Chicago administers a Safe Space Program offering workshops and other resources for "developing an ally network."[5] Other people, even within a single institution, assert that safe spaces coddle students and stifle the expression of "scary" ideas (Shulevitz 2016). The University of Chicago Dean of Students issued a letter to the incoming class of 2020 declaring, "Our commitment to academic freedom means...we do not condone the creation of intellectual 'safe spaces' where individuals can retreat from ideas and perspectives at odds with their own." Countless education experts have weighed in on the safe-spaces debate (e.g., Campbell and Manning 2018; Leonardo and Porter 2010; Groen 2008), but their theses are almost always editorial or theoretical rather than empirical. The existing literature therefore has limited potential to inform social decision-making and constructive education reform. Beyond its core methodological contribution, this experiment will offer clearer recommendations for designing and implementing safe-space policies.

2 Experimental Design

When subjective and personal norms conflict, psychologists theorize that the most salient or "focal" norm will have a the strongest influence on a subject's actions (Cialdini and Trost 1998, 162). This implies that researchers can alter the salience of norms so as to maximize the chance of honest responses on questionnaires. I will test that implication by manipulating the messages respondents receive before answering questions about socially sensitive behaviors and attitudes. Specifically, I will test whether reassuring respondents of a safe space makes them more likely to give "undesirable" or "awkward" but honest answers. This *framing* treatment will encourage respondents to think about their answers as unlikely to elicit negative reactions. Framing, as Popkin (1994, 84) defines it, occurs "whenever there is more than one way to think about a subject," and when certain information is "brought to the fore." Framing experiments are common in political science and economics (e.g., Kühberger 1998; Slothuus and De Vreese 2010), but none to date has

<hr>

[5]https://lgbtq.uchicago.edu/page/lgbtq-safe-space

tested for the effects of safe spaces.

Earlier approaches to creating safe spaces in a research design are not practicable in medium-*n* or large-*n* surveys. For instance, Stoudt (2007) had to "work hard" in a study of school violence to make student subjects trust him, by interacting repeatedly with individual students. Kamberelis and Dimitriadis (2011) gave female participants in focus groups time to get to know each other, share their experiences, and "reclaim their humanity." Besides being difficult to replicate with larger samples, it is unclear whether these precautions caused subjects to be more candid, because there was no control group in either study, and because subjects who agreed to participate might have differed systematically from subjects who did not. The experiment described below circumvents these problems by randomizing the treatment of safe spaces across subjects.

2.1 Intervention

Questionnaires are appealing because they allow an investigator to collect dozens of responses from each of hundreds or thousands of people. Nurturing trust through time-intensive interactions with every subject would undermine this efficiency. Therefore, the experimental treatment consists of a short preamble at the start of a questionnaire, with periodic reminders, reassuring respondents that they can express themselves in a safe environment. It is derived from a clause that the Center for Teaching and Learning at a U.S. university recommended as a model for instructors to include in their class syllabi. Borrowing language from a real syllabus lends external validity to my analysis, although I apply the clause to a non-classroom setting. Below are the original syllabus clause and the treatment preamble. I do not use the exact term "safe space" in the treatment preamble, in order to separate the effects of safe spaces from the effects of politically charged question wording.

> **Syllabus clause:** "My office is a safe space, and I am friendly toward all groups of students including those who identify as LGBTQ+. I want you to know that you can feel safe from judgmental attitudes when interacting with me."

> **Treatment preamble:** "This survey will ask you to share your views and basic infor-

mation about yourself. Your responses will remain anonymous. **This is an environment in which you should feel comfortable expressing yourself honestly and participating fully, without fear of attack, ridicule, or judgment. The researchers will respect all responses, including uncommon positions.** By proceeding, you agree to participate."

The bold text represents the rhetorical emphasis that instructors typically place on safe-space language. Although this might raise concerns about experimenter demand bias, research suggests the risks of such bias are minimal (de Quidt and Roth 2018).

The control version of the survey will include only a reassurance of anonymity and low risk:

Control preamble: "This survey will ask you to share your views and basic information about yourself. Your responses will remain anonymous. By proceeding, you agree to participate."

One common objection to survey experiments is that any significant treatment effects have an unknown but presumably short duration (Gaines, Kuklinski and J. 2007, 5). In the context of safe-space priming, however, the usual goal of the educator (or researcher) issuing the safe-space reassurance is to make students (or respondents) feel comfortable expressing themselves honestly in the moment. Instilling a longer-term feeling of safety is a secondary concern.

2.2   Sample

The questionnaire and treatment will be administered electronically to adult U.S. subjects using a professional survey firm. The total sample size will be approximately 2000 people (1000 each in the treatment and control samples), which is more than sufficient to detect an effect size of .12 [a "small" effect size according to Cohen (1988)] with a probability of .8 (one-sided test, $\alpha = .05$). Figure 1 illustrates how the necessary sample size changes with statistical power.
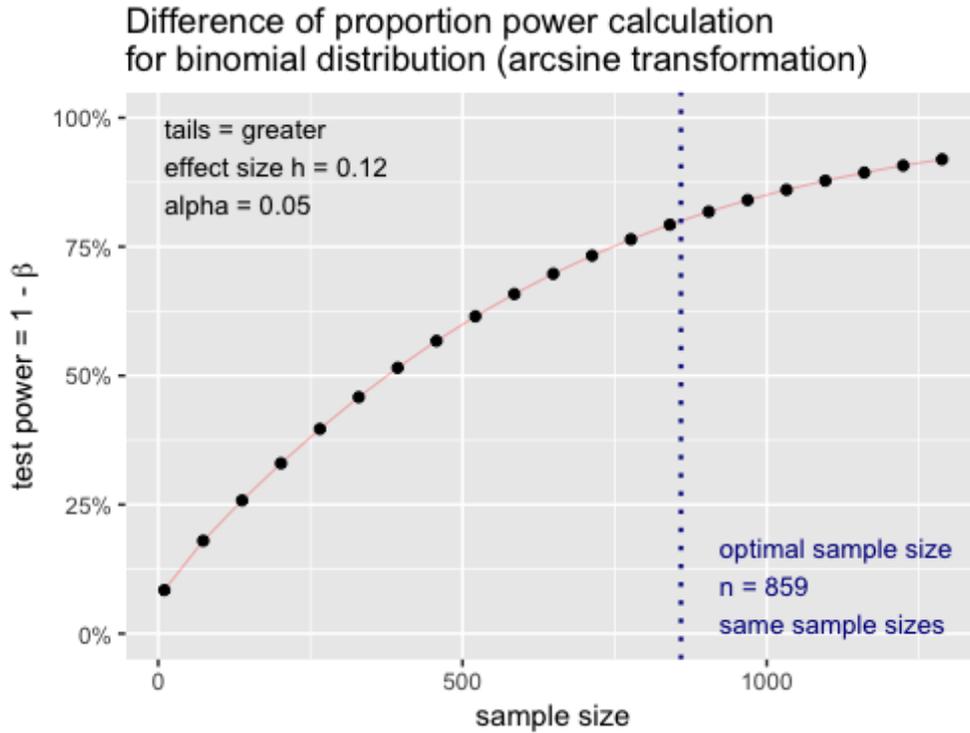
Figure 1: Power Calculation

2.3 Random Assignment of Treatment

Treatments will be randomized using the "Percent Branch" function of the software SurveyGizmo. For each respondent, this function randomly chooses a number between 1 and 100 and then uses it to assign the respondent to one of two branches corresponding to the two preambles specified in Section 2.1, where each branch has a 50 percent chance of being assigned. For example, if the randomly chosen number is 27, the function will assign the respondent to branch/preamble 1; if it is 57, the function will assign the respondent to branch/preamble 2. I will use a one-sided difference-in-means test ($\alpha = .05$) to assess whether sensitivity bias is lower on average for respondents randomly assigned to the treatment condition than it is for respondents assigned to control.

3 Outcomes

The outcome of interest is sensitivity bias. The intuitive way to measure it is by comparing respondents' survey answers to their observed behavior. For instance, Karp and Brockington (2005) compared self-reported voter turnout to government records of individual voter turnout. However, it is unusual to have access to such records of respondent behavior. Indeed, Karp and Brockington were able to procure turnout data of varying quality from only five countries. For their study of voter over-reporting in Benin, Adida et al. (2018) did not obtain individual-level data at all, instead aggregating polling station data up to the village level. In addition to data availability problems, studies that rely on administrative records are not immune to reporting bias on the part of government officials. And administrative records cannot let researchers study sensitivity bias in survey answers pertaining to *attitudes*, which are never directly observable. This is why, "as an empirical matter, it is unknown whether measures of attitudes or behaviors face a greater threat of sensitivity bias" (Blair, Coppock and Moor 2018, 8). To overcome these limitations of the literature, I measure sensitivity bias at the individual level in two ways: 1) the Marlowe-Crowne Social Desirability Scale (which addresses respondent *behaviors*); and 2) the difference between direct and combined (list and direct) measures of respondent *attitudes*.

3.1 Marlowe-Crowne Social Desirability Scale

The Marlowe-Crowne Social Desirability Scale (MCSDS) (Crowne and Marlowe 1960) is the most common and a widely validated scale for measuring social desirability bias in psychology and sociology (Nederhof 1985; Lambert, Arbuckle and Holden 2016), although few political scientists use it.[6] It consists of 33 yes-or-no questions about behaviors that are "culturally sanctioned and approved, but which are improbable of occurrence," such as "I have never intensely disliked anyone" and "Before voting I thoroughly investigate the qualifications of all the candidates." Respondents

---

[6]There are exceptions, including Bishop, Tuchfarber and Oldendick (1986).

who give unlikely answers to more questions score higher on social desirability bias. Because safe spaces theoretically enhance people's willingness to share all kinds of sensitive truths about themselves (Holley and Steiner 2005), I expect the components of the MCSDS to respond to the safe-space treatment.

Following Ballard (1992), Theriault and Holmberg (1998), and Strahan and Gerbasi (1972), I use a shortened version of the MCSDS to keep the entire questionnaire brief and to "avoid arousing evaluation apprehension." Scholars have shown shortened formats to be reliable (e.g., Lambert, Arbuckle and Holden 2016; Ray 1984; Sârbescu, Costea and Rusu 2012; Strahan and Gerbasi 1972), despite some arguments to the contrary (e.g., Barger 2002; Leite and Beretvas 2005). The MCSDS is controversial, and shortened versions even more so. However, recent analyses demonstrate that the MCSDS outperforms more "advanced" scales such as the Impression Management (IM) scale of the Balanced Inventory of Desirable Responding (BIDR): "This result is counterintuitive because the IM scale and its parent BIDR are seen as the new gold standard for measuring socially desirable responding. Further, the MCSDS is commonly considered to be outdated due to its age and its true/false response format" (Lambert, Arbuckle and Holden 2016, 85). An analysis by Loo and Loewen (2004) specifically supports the use of shortened versions of the MCSDS, identifying Ballard's (1992) Scale 1 as the superior choice. This is the version I use in my survey (Section 8).

A disadvantage of the MCSDS is that it was developed to measure social desirability bias, which is only one facet of sensitivity bias (see Section 1). I therefore use a second measure of sensitivity bias, described below.

3.2 Difference between Direct and Combined (List and Direct) Measures

As explained in Section 1, list experiments are a popular tool for inferring actual (versus self-reported) characteristics of survey respondents, thus helping to correct for sensitivity bias. But they can also serve as a *measure* of sensitivity bias if one compares the estimated rate of socially undesirable traits from a list experiment to the rate from self-reports in response to direct questions.

For example, Gonzalez-Ocantos et al. (2012) found that the difference between list-experiment and self-reported rates of receiving a gift or favor in exchange for a vote was 13 percentage points for Nicaraguan men and 31 percentage points for Nicaraguan women, suggesting that women had higher rates of sensitivity bias.

In the present study on U.S. subjects, I use a similar experimental approach to measure sensitivity bias in respondent reports of believing that African-American athletes are disrespectful to take a knee during the national anthem. Previous studies have revealed substantial bias in responses to questions on racial issues (Blair, Coppock and Moor 2018), making this an appropriate topic for measuring variations in bias across subjects who receive or do not receive safe-space priming. I calculate a measure of bias from three survey questions: one asking respondents to report how many attitudes they hold from a list that includes opposition to taking a knee, the same question but with a list of attitudes that excludes opposition to taking a knee, and a question asking directly whether respondents oppose taking a knee (Section 8). All respondents will receive the direct question,[7] while each respondent will receive only either the longer (treatment) or the shorter (control) version of the list question, assigned at random. Following Eady (2017), I will separate the list and direct questions by unrelated questions and randomize the order of list and direct questions, so as to reduce the chance that treatment status affects answers to the direct question. Systematic differences in the means of reported list items across subjects seeing the treatment versus the control list will suggest the true prevalence of opposition to taking a knee; the direct question will indicate the biased prevalence that respondents feel safe divulging. For example, if respondents who see the treatment list report 2.5 items on average and respondents who see the control list report 2.3 items on average, we can infer that 20 percent of respondents (2.5 - 2.3 = 0.2) think taking a knee is disrespectful. The difference between that inferred value and the percent of respondents who self-report opposition to taking a knee can then provide a measure of sensitivity bias. For example, if 5

_____

[7]See Aronow et al. (2015, 45) on why it is better to ask the direct question whenever feasible, as opposed to posing the direct question only to a subset of respondents, as some studies do.

percent of respondents say directly that they oppose taking a knee, the measure of sensitivity bias would be 15 percentage points (20 - 5). By replicating this list experiment under each randomized preamble condition, I can compare sensitivity bias between the main treatment group (safe-space preamble) and control group (no safe-space preamble).

Some scholars have raised concerns that list experiments are poor measures of true beliefs and behaviors due to the method's imprecision (Blair, Coppock and Moor 2018; Eady 2017; Chou, Imai and Rosenfeld 2017). For a more accurate measure of sincere attitudes, I will analyze the survey data using the R function *combinedListDirect* (Blair et al. 2016), which implements an estimator developed by Aronow et al. (2015) for combining information from list and direct questions in order to exploit both the higher precision of direct questions and the lower bias of list questions.[8] Analogous to other differences-in-differences methods, I will then compare discrepancies between the combined (list and direct) estimates of racial attitudes and the direct estimates, both for the group that receives the safe-space treatment and for the group that does not. Table 1 summarizes the assignment of subjects to conditions in the list experiment nested within the main experiment on the effects of safe-space language. 1000 subjects is lower than the threshold at which it is typically possible to detect sensitivity bias using list experiments (Blair, Coppock and Moor 2018, 17). My detection ability will consequently be slightly under-powered, though my sample will still be larger than in many other list experiments (e.g., Gonzalez-Ocantos et al. 2012; Kuklinski et al. 1997).

Safe-Space Preamble		Control Preamble	
Treatment List	Control List	Treatment List	Control List
N ≈ 500	N ≈ 500	N ≈ 500	N ≈ 500

Table 1: Assignment of Subjects to Experimental Conditions

All groups also receive direct question.

---

[8]This approach reduces variability in list experiments between 12 and 50 percent, which is comparable to other techniques (Blair, Coppock and Moor 2018, 20). One can reduce variability up to 88 percent by incorporating auxiliary information that hints at the true prevalence of a sensitive trait (Chou, Imai and Rosenfeld 2017). However, auxiliary information is rarely available (Blair, Coppock and Moor 2018, 20).

Neither of the above measures is perfect, but scholars have endorsed both as reasonable proxies for directly observing respondent behavior. One potential objection to using the Marlowe-Crowne scale is that subjects might have different pre-treatment susceptibility to sensitivity bias. However, random assignment should balance susceptibility to sensitivity bias across treatment and control groups; balance checks using other covariates will indicate whether random assignment "worked." A potential objection to using the combined estimator is that it will remain biased if subjects vary systematically in having the sensitive trait (opposition to taking a knee). Yet, I do not expect the treatment to affect whether people *have* the trait, only whether they admit to having it. The combined estimator also assumes, reasonably, that a direct admission of having the sensitive trait reveals that the respondent indeed has the sensitive trait. Ultimately, using two different measures of sensitivity bias will permit robustness checks, producing suggestive evidence of patterns in sensitivity bias across treatment and control groups.

4 Covariates

Although random assignment should produce unbiased estimates of treatment effects without controlling for other variables, adjusting for covariates can give more precise estimates should there be any incidental imbalance between treatment and control groups. In multivariate regressions of the Marlowe-Crowne Social Desirability Scale (which is measured at the individual level), covariates will include a respondent's gender, age, race, education level, income level, religiosity, and party identification, and who the respondent thinks commissioned the study. I will report results with and without controls. It is possible that the safe-space treatment could affect respondents' willingness to answer even basic demographic questions, which could introduce post-treatment bias into statistical estimates. It will therefore be important to take the models that include covariates with a grain of salt.

5 Hypotheses

5.1 Main Effects

- Hypothesis 1: Reassurance of a safe space will reduce the Marlowe-Crowne score.

- Hypothesis 2: Reassurance of a safe space will reduce the difference between combined and direct measures of belief that African-American athletes are disrespectful to take a knee during the national anthem.

5.2 Heterogeneous Treatment Effects

There are compelling reasons to test for heterogeneous treatment effects across gender and racial sub-groups. Studies have consistently shown women to display more sensitivity bias than men (Krumpal 2013; Bernardi and Guptill 2008; Guadagno and Cialdini 2007; Hebert et al. 1997; Schoderbek and Deshpande 1996; Cohen, Pant and Sharp 1998; Press and Townsley 1998; Gonzalez-Ocantos et al. 2012). By one estimate, gender accounts for 41 percent of all variation in socially desirable responding (Bernardi 2006). Other studies have revealed greater sensitivity bias among nondominant ethnic or racial communities (Odendaal 2015; Johnson and Van de Vijver 2015). Some authors insinuate that these differences are primordial "cultural orientations" (Lalwani, Shavitt and Johnson 2006) or reflect underlying disparities in cognitive ability (Odendaal 2015). Yet, it is plausible that socially desirable responding is actually a mechanism for marginalized people to cope with the stigma they face for appearing "bitchy," "aggressive," "ungrateful," and so forth. Chwe (2014, 29) documents how women and African-American slaves historically used strategic thinking to manage risk in a world stacked against them. Anticipating the risk of social sanctioning, respondents from nondominant groups might alter their sincere answers to be more equivocal or socially desirable, having learned such adaptation through life experiences. For instance, women who strive to be likeable at work receive better performance reviews than those

who do not, even though likability has no effect on the reviews of male employees (Kipnis and Schmidt 1988). Hochschild (2012) famously observed that women tend to fill jobs, such as flight attendant and school teacher, that require "emotional labor," or "the process by which workers are expected to manage their feelings in accordance with organizationally defined rules and guidelines" (Wharton 2009, 149). As part of the work of survey-taking,[9] social desirability bias in particular is what Hochschild terms "surface acting"—the public display of dispositions in line with other people's expectations.[10] The very notions of "desirability" and "faking it" evoke gender norms. Because women feel acute pressure to conform to social expectations, they may be especially responsive to safe spaces (Harris 2005). The same may be true of people from marginalized racial groups. In the U.S. context these are nonwhites, though scholars note that black Americans have faced pronounced discrimination, even within other communities of color (Sexton 2008).

- Hypothesis 3: Reassurance of a safe space will reduce sensitivity bias more if the respondent is female as opposed to male.

- Hypothesis 4: Reassurance of a safe space will reduce sensitivity bias more if the respondent is nonwhite as opposed to white.

- Hypothesis 5: Reassurance of a safe space will reduce sensitivity bias more if the respondent is black as opposed to nonblack.

[9]The concept of survey-taking-as-labor occasionally arises in debates about compensating respondents (Ripley et al. 2010), but rarely do such debates acknowledge *emotional* labor (as opposed to time spent).

[10]Surface acting correlates with burnout, family conflict, and other social-psychological problems (Noor and Zainuddin 2011), so reducing it has implications for respondents' mental health as well as for the validity of questionnaire-based research.

6 Analysis

The regression models for analyzing average treatment effects will take the following form:

$$E(Y_i) = \beta_0 + \beta_1 T_i + \mathbf{X_i}$$

where Y_i is a measure of sensitivity bias, T_i is a dichotomous variable taking a value of 1 if subject i received the safe-space treatment and 0 otherwise, and $\mathbf{X_i}$ is a vector of covariates. As mentioned previously, I will report results with and without controls (see Section 4).

When evaluating heterogeneous treatment effects, I will use the following interaction equation:

$$E(Y_i) = \beta_0 + \beta_1 T_i + \beta_2 M_i + \beta_3 T_i \times M_i + \mathbf{X_i}$$

where M_i is the moderator of interest.

7 Corrections

7.1 Attrition and Noncompliance

I anticipate low attrition and noncompliance, as the sample will consist of a high-quality online panel. Nevertheless, I will use intention-to-treat analysis wherein I include every subject who is randomly assigned to treatment or control, ignoring attrition or noncompliance that occurs after random assignment. This will produce conservative estimates of treatment effects (Gupta 2011).

7.2 Outcomes with Limited Variation

I will drop from the analysis any questionnaire items for which 95 percent of observations have the same value.

7.3 Multiple Hypothesis Testing

When testing multiple hypotheses, there is a possibility that some tests will have *p*-values less than .05 purely by chance. To limit the risk of Type I errors, I will apply the Benjamini and Hochberg (1995) procedure, with a false discovery rate of .05 (meaning up to 5 percent of results that appear positive are truly negative). I prefer this to the classic Bonferroni correction because it strikes a good balance between preventing Type I and Type II errors.

7.4 Design Effects

As noted in Section 1, the wording of a survey question might bias responses. One way I mitigate this bias is by including decoy list questions and non-list questions to distract respondents from the questions used to measure sensitivity bias (see survey instrument in Section 8). Budget permitting, I will also repeat the experiment with a differently worded treatment preamble, derived from a different syllabus clause, to test the robustness of results:

> **Alternative syllabus clause:** "The classroom is a special environment in which students and faculty come together to promote learning and growth. It is essential to this learning environment that respect for the rights of others seeking to learn, respect for the professionalism of the instructor, and the general goals of academic freedom are maintained. Differences of viewpoint or concerns should be expressed in terms which are supportive of the learning process, creating an environment in which students and faculty may learn to reason with clarity and compassion, to share of themselves without losing their identities, and to develop an understanding of the community in which they live. Student conduct which disrupts the learning process shall not be tolerated and may lead to disciplinary action and/or removal from class."

> **Alternative treatment preamble:** "This survey will ask you to share your views and basic information about yourself. Your responses will remain anonymous. This survey is a special environment in which respondents have a chance to promote scholarly learning. To ensure an effective learning environment, the researchers will analyze all responses with respect and professionalism. Any differences of viewpoint or concerns that you share of yourself are welcome. The aim of this study is to develop an

understanding of society without any judgment of how respondents choose to express themselves. By proceeding, you agree to participate."

7.5 External Validity

Because educators generally cultivate safe spaces with students in person, it is fitting to consider whether the safe-space treatment reduces sensitivity bias when interviews occur in person and with students. To test the external validity of the main, online-based results in a more realistic setting, I will replicate (budget permitting) the survey experiment on a panel of American university students, with enumerators interviewing the students face-to-face. Statistical analyses of the resulting data will include enumerator fixed effects to control for possible enumerator bias, resulting in the following variant of the main regression model specified in Section 7:

$$E(Y_i) = \beta_0 + \beta_1 T_i + \mathbf{X_i} + \mathbf{u_j}$$

where $\mathbf{u_j}$ denotes fixed effects for each enumerator j.

The interaction model becomes the following after adding enumerator fixed effects:

$$E(Y_i) = \beta_0 + \beta_1 T_i + \beta_2 M_i + \beta_3 T_i \times M_i + \mathbf{X_i} + \mathbf{u_j}$$

For practical reasons, I will aim for a sample of at least 500 students, resulting in an underpowered but hard test of my hypotheses. Because politically conservative students may be unwilling to give candid survey responses in a politically progressive campus environment or vice versa, the treatment preamble will explicitly state that both liberal and conservative viewpoints are welcome:

> **Treatment preamble (student version):** "This survey is an environment in which you should feel comfortable expressing yourself honestly and participating fully, without fear of attack, ridicule, or judgment. The researchers will respect all groups of respon-

dents including those who identify with underrepresented positions, whether liberal or conservative.[11] Your responses to this survey will remain strictly anonymous."

8 Survey Instrument

Items are organized here by module, not their final order in the questionnaire. Details on final order noted below. Respondents in the treatment group will see the safe-space language at the start of the survey and after the Marlowe-Crowne scale.

Module 1: Short Marlowe-Crowne Social Desirability Bias Scale[12] (direction of keying in parentheses)

1. (F) I sometimes feel resentful when I don't get my way.

2. (F) On a few occasions, I have given up doing something because I thought too little of my ability.

3. (F) There have been times when I felt like rebelling against people in authority even though I knew they were right.

4. (T) No matter who I'm talking to, I'm always a good listener.

5. (F) I can remember "playing sick" to get out of something.

6. (F) There have been occasions when I took advantage of someone.

7. (T) I'm always willing to admit it when I make a mistake.

8. (F) I sometimes try to get even rather than forgive and forget.

9. (T) When I don't know something I don't at all mind admitting it.

10. (F) I am sometimes irritated by people who ask favors of me.

11. (T) I have never deliberately said something that hurt someone's feelings.

Module 2: Racial Attitudes

Respondents will see the direct and list versions of the racial attitudes question in random order, to avoid question-order effects. (Aronow et al. 2015).

In designing the lists below, I had two goals. First, I wanted to avoid bias from "ceiling effects" that can result when lists include too many low-variance items (thus reducing privacy protection for respondents) (Glynn 2013). However, I also wanted to reduce variance in responses for the quality

---

[11]Order of "liberal" and "conservative" will be randomized.

[12]Ballard (1992) Scale 1

of the list experiment. I used a method suggested by Glynn (2013) to accomplish both goals, which involves introducing negative correlation within each list. Namely, respondents who support gay marriage are unlikely to also support prayer in schools. I also include a high-endorsement item ("American politics is too polarized") along with a low-endorsement item ("The voting age should be raised to 25").

1. (Control list) Below is a list of political beliefs. Please indicate how many of these beliefs you hold. Do not indicate which ones, just how many.

   (a) The voting age should be raised to 25.

   (b) People of the same sex should be able to legally marry.

   (c) Prayer should be allowed in public schools.

   (d) American politics is too polarized.

2. (Treatment list) Below is a list of political beliefs. Please indicate how many of these beliefs you hold. Do not indicate which ones, just how many.

   (a) The voting age should be raised to 25.

   (b) People of the same sex should be able to legally marry.

   (c) It is disrespectful for African-American athletes to take a knee during the national anthem.

   (d) Prayer should be allowed in public schools.

   (e) American politics is too polarized.

3. (Direct question, to appear long after or before list question. To facilitate comparison with the list experiment, I will dichotomize answers into "agree" and "do not agree.") To what extent do you agree or disagree with the following statement? "It is disrespectful for African-American athletes to take a knee during the national anthem."

   (a) Completely agree

   (b) Somewhat agree

   (c) Neither agree nor disagree

   (d) Somewhat disagree

   (e) Completely disagree

Attention Screener

1. Which of the following activities do you think has the greatest influence on politics?

   (a) Voting

   (b) Joining a demonstration

(c) Contacting a public official

(d) Donating money to a campaign or cause

2. The previous question asked you to choose from several activities. Which of the following was on that list? (Following Eady (2017), I will eliminate from the results any subjects who do not answer "joining a demonstration.")

(a) Volunteering for an organization

(b) Joining a demonstration

(c) Writing an op-ed

(d) Raising awareness on social media

Decoy Questions

The main purpose of these questions is to prevent respondents from identifying racial attitudes as the focus of the main experiment, but they also permit a second list experiment to cross-validate the main results. Following Glynn (2013), I will randomize the order of lists and the order of direct questions.

1. (Alternative control list) Below is a list of political beliefs. Please indicate how many of these beliefs you hold. Do not indicate which ones, just how many.

(a) More young people should run for office.

(b) There should be a gender quota in the U.S. Congress.

(c) The U.S. should intervene less in other countries.

(d) The tax on gasoline should double.

2. (Alternative treatment list) Below is a list of political beliefs. Please indicate how many of these beliefs you hold. Do not indicate which ones, just how many.

(a) More young people should run for office.

(b) There should be a gender quota in the U.S. Congress.

(c) Men are better suited to be president than women.

(d) The U.S. should intervene less in other countries.

(e) The tax on gasoline should double.

3. (Alternative direct question) To what extent do you agree or disagree with the following statement? "Men are better suited to be president than women."

(a) Completely agree

(b) Somewhat agree

(c) Neither agree nor disagree

(d) Somewhat disagree

(e) Completely disagree

4. To what extent do you agree or disagree with the following statement? "Marijuana should be legalized nationwide."

(a) Completely agree

(b) Somewhat agree

(c) Neither agree nor disagree

(d) Somewhat disagree

(e) Completely disagree

5. Please rank the following policy issues in order of importance to you.

(a) Economy

(b) Justice/courts

(c) Foreign affairs

(d) Education

(e) Law enforcement

(f) Environment

Module 3: Demographics and Manipulation Check[13]

1. What is your gender?

(a) Male

(b) Female

(c) Other

2. In what year were you born?

3. Are you Spanish, Hispanic, or Latino?

(a) Yes

(b) No

4. Please choose one or more races that you consider yourself to be:

(a) White

---

[13]I mimic ANES wording to be able to compare my sample to the nationally-representative ANES sample.

(b) Black or African-American

(c) American Indian or Alaska Native

(d) Asian

(e) Native Hawaiian or other Pacific Islander

(f) Other [specify]

5. What is the highest level of school you have completed or the highest degree you have received?

(a) Less than 1st grade

(b) 1st, 2nd, 3rd, or 4th grade

(c) 5th or 6th grade

(d) 7th or 8th grade

(e) 9th grade

(f) 10th grade

(g) 11th grade

(h) 12th grade no diploma

(i) High school graduate - High school diploma or equivalent (For example: GED)

(j) Some college but no degree

(k) Associate degree in college - Occupational/vocational program

(l) Associate degree in college - Academic program

(m) Bachelor's degree (For example: BA, AB, BS)

(n) Master's degree (For example: MA, MS, MEng, MEd, MSW, MBA)

(o) Professional school degree (For example: MD, DDS, DVM, LLB, JD)

(p) Doctorate degree (For example: PhD, EdD)

(q) Other [specify]

6. What was the total income in 2018 of all the members living in your household? Your best guess is fine. [enter income]

7. How much guidance does religion provide in your day-to-day living?

(a) None

(b) Some

(c) Quite a bit

(d) A great deal

8. Generally speaking, how do you identify in terms of political party? (Order of parties will be randomized.)

   (a) No preference

   (b) Democrat

   (c) Republican

   (d) Independent

   (e) Other party [specify]

9. Who do you think commissioned this survey?

   (a) Government

   (b) University or school

   (c) Research institution (besides universities or schools)

   (d) Political party

   (e) Nonprofit group

   (f) Private company

   (g) Media

   (h) Other [specify]

10. (Manipulation check) Which of the following were the instructions given at the beginning of the survey?

   (a) This survey will ask you to share your views and basic information about yourself.

   (b) This survey will ask you to share your views and basic information about yourself. Your responses will remain anonymous. By proceeding, you agree to participate.

   (c) This survey will ask you to share your views and basic information about yourself. Your responses will remain anonymous. This is an environment in which you should feel comfortable expressing yourself honestly and participating fully, without fear of attack, ridicule, or judgment. The researchers will respect all groups of respondents including those who identify with uncommon positions. By proceeding, you agree to participate.

References

Adida, Claire L., Jessica Gottlieb, Eric Kramon and Gwyneth H. McClendon. 2018. "Response Bias in Survey Measures of Voter Behavior: Implications for Measurement and Inference." Unpublished manuscript.

Adida, Claire L., Karen E. Ferree, Daniel N. Posner and Amanda Lea Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49(12):1630–1660.

Arao, Brian and Kristi Clemens. 2013. From Safe Spaces to Brave Spaces. In *The Art of Effective Facilitation: Reflections from Social Justice Educators*, ed. Lisa M. Landreman. Sterling, VA: Stylus pp. 135–150.

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

Ballard, Rebecca. 1992. "Short Forms of the Marlowe-Crowne Social Desirability Scale." *Psychological Reports* 71:1155–1160.

Barger, Steven D. 2002. "The Marlowe-Crowne Affair: Short Forms, Psychometric Structure, and Social Desirability." *Journal of Personality Assessment* 79(2):286–305.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society* 57(1):289–300.

Berinski, Adam. 2004. "Can We Talk? Self-Presentation and the Survey Response." *Political Psychology* 25(4):643–659.

Bernardi, Richard A. 2006. "Associations between Hofstede's Cultural Constructs and Social Desirability Response Bias." *Journal of Business Ethics* 65(1):43–53.

Bernardi, Richard A. and Steven T. Guptill. 2008. "Social Desirability Response Bias, Gender, and Factors Influencing Organizational Commitment: An International Study." *Journal of Business Ethics* 81(4):797–809.

Bishop, George F., Alfred J. Tuchfarber and Robert W. Oldendick. 1986. "Opinions on Fictitious Issues: The Pressure to Answer Survey Questions." *Public Opinion Quarterly* 50(2):240–250.

Blair, Graeme, Alexander Coppock and Margaret Moor. 2018. "When to Worry About Sensitivity Bias: Evidence from 30 Years of List Experiments." Unpublished manuscript.

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

Blair, Graeme, Kosuke Imai, Bethany Park, Alexander Coppock and Winston Chou. 2016. "list: Multivariate Statistical Analysis for the Item Count Technique." R package version 8.3.

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, Kosuke Imai and Yang-Yang Zhou. 2015. "Design and Analysis of the Randomized Response Technique." *Journal of the American Statistical Association* 110(511):1304–1319.

Blaydes, Lisa and Rachel M. Gillum. 2013. "Religiosity-of-Interviewer Effects: Assessing the Impact of Veiled Enumerators on Survey Response in Egypt." *Politics & Religion* 6(3).

Campbell, Bradley and Jason Manning. 2018. *The Rise of Victimhood Culture: Microaggressions, Safe Spaces, and the New Culture Wars*. Cham, Switzerland: Palgrave Macmillan.

Chou, Winston, Kosuke Imai and Bryn Rosenfeld. 2017. "Sensitive Survey Questions with Auxiliary Information." *Sociological Methods & Research* pp. 1–37.

Chwe, Michael Suk-Young. 2014. *Jane Austen, Game Theorist*. Princeton: Princeton University Press.

Cialdini, Robert B. and Melanie R. Trost. 1998. Social Influence: Social Norms, Conformity and Compliance. In *The Handbook of Social Psychology*, ed. Susan T. Fiske, Daniel T. Gilbert and Gardner Lindzey. New York: McGraw-Hill pp. 151–192.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, Jeffrey R., Laurie W. Pant and David J. Sharp. 1998. "The Effect of Gender and Academic Discipline Diversity on the Ethical Evaluations, Ethical Intentions and Ethical Orientation of Potential Public Accounting Recruits." *Accounting Horizons* 12(3):250–270.

Crowne, Douglas P. and David Marlowe. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24(4):349–354.

de Quidt, Jonathan adn Haushofer, Johannes and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108(11):3266–3302.

Downes, Sophie. 2016. "Trigger Warnings, Safe Spaces and Free Speech, Too." *New York Times*. September 10, 2016 .

Eady, Gregory. 2017. "The Statistical Analysis of Misreporting on Sensitive Survey Questions." *Political Analysis* 25(2):241–259.

Felson, Richard B. and Paul-Philippe Paré. 2005. "The Reporting of Domestic Violence and Sexual Assault by Nonstrangers to the Police." *Journal of Marriage and Family* 67(3):597–610.

Gaines, Brian J., James H. Kuklinski and Quirk Paul J. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15(1):1–20.

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

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet De Jonge, Carlos Meléndez, Javier Osorio and David W Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.

Groen, Janet. 2008. "Paradoxical Tensions in Creating a Teaching and Learning Space within a Graduate Education Course on Spirituality." *Teaching in Higher Education* 13(2):193–204.

Guadagno, Rosanna E. and Robert B. Cialdini. 2007. "Gender Differences in Impression Management in Organizations: A Qualitative Review." *Sex Roles* 56(7-8):483–494.

Gupta, Sandeep K. 2011. "Intention-to-Treat Concept: A Review." *Perspectives in Clinical Research* 2(3):109–112.

Hainmueller, Jens and Michael J. Hiscox. 2010. "Attitudes toward Highly Skilled and Low-Skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review* 104(1):61–84.

Harris, Anita. 2005. "Discourses of Desire as Governmentality: Young Women, Sexuality and the Significance of Safe Spaces." *Feminism & Psychology* 15(1):39–43.

Hatchett, Shirley and Howard Schuman. 1975. "White Respondents and Race-of-Interviewer Effects." *Public Opinion Quarterly* 39(4):523–528.

Hebert, James R., Yunsheng Ma, Lynn Clemow, Ira S. Ockene, Gordon Saperia, Edward J. Stanek III, Philip A. Merriam and Judith K. Ockene. 1997. "Gender Differences in Social Desirability and Social Approval Bias in Dietary Self-Report." *American Journal of Epidemiology* 146(12):1046–1055.

Hochschild, Arlie Russell. 2012. *The Managed Heart*. Berkeley: University of California Press.

Holbrook, Allyson L. and Jon A. Krosnick. 2009. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1):37–67.

Holley, Lynn C. and Sue Steiner. 2005. "Safe Space: Student Perspectives on Classroom Environment." *Journal of Social Work Education* 41(1):49–64.

Janis, Irving L. 1971. "Groupthink." *Psychology Today* 5(6):84–90.

Johnson, Timothy P. and Fons J. R. Van de Vijver. 2015. Social Desirability in Cross-Cultural Research. In *Cross-Cultural Survey Methods*, ed. Janet A. Harkness, Fons J. R. Van de Vijver and Peter Ph. Mohler. Hoboken, NJ: John Wiley & Sons pp. 193–209.

Kamberelis, George and Greg Dimitriadis. 2011. Focus Groups: Contingent Articulations of Pedagogy, Politics, Inquiry. In *The SAGE Handbook of Qualitative Research*, ed. Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE pp. 545–562.

Karp, Jeffrey A. and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *Journal of politics* 67(3):825–840.

Katz, Daniel. 1942. "Do Interviewers Bias Poll Results?" *Public Opinion Quarterly* 6(2):248–268.

Kipnis, David and Stuart M Schmidt. 1988. "Upward-Influence Styles: Relationship with Performance Evaluations, Salary, and Stress." *Administrative Science Quarterly* pp. 528–542.

Krumpal, Ivar. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality & Quantity* 47(4):2025–2047.

Kühberger, Anton. 1998. "The Influence of Framing on Risky Decisions: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 75(1):23–55.

Kuklinski, James H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence and Barbara Mellers. 1997. "Racial Prejudice and Attitudes toward Affirmative Action." *American Journal of Political Science* pp. 402–419.

Lalwani, Ashock K., Sharon Shavitt and Timothy Johnson. 2006. "What Is the Relation Between Cultural Orientation and Socially Desirable Responding?" *Journal of Personality and Social Psychology* .

Lambert, Christine E., Spencer A. Arbuckle and Ronald R. Holden. 2016. "The Marlowe-Crowne Social Desirability Scale Outperforms the BIDR Impression Management Scale for Identifying Fakers." *Journal of Research in Personality* 61:80–86.

Leite, Walter L. and S. Natasha Beretvas. 2005. "Validation of Scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding." *Educational and Psychological Measurement* 65(1):140–154.

Leonardo, Zeus and Ronald K. Porter. 2010. "Pedagogy of Fear: Toward a Fanonian Theory of 'Safety' in Race Dialogue." *Race Ethnicity and Education* 13(2):139–157.

Loo, Robert and Pamela Loewen. 2004. "Confirmatory Factor Analyses of Scores from Full and Short Versions of the Marlowe-Crowne Social Desirability Scale." *Journal of Applied Social Psychology* 34(11):2343–2352.

McClendon, Gwyneth H. 2018. *Envy in Politics*. Princeton: Princeton University Press.

Nederhof, Anton J. 1985. "Methods of Coping with Social Desirability Bias: A Review." *European Journal of Social Psychology* 15(3):263–280.

Noor, Noraini M. and Masyitah Zainuddin. 2011. "Emotional Labor and Burnout among Female Teachers: Work–Family Conflict as Mediator." *Asian Journal of Social Psychology* 14(4):283–293.

Odendaal, Aletta. 2015. "Cross-Cultural Differences in Social Desirability Scales: Influence of Cognitive Ability." *SA Journal of Industrial Psychology* 41(1):1–13.

Popkin, Samuel L. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.

Press, Julie E. and Eleanor Townsley. 1998. "Wives' and Husbands' Housework Reporting: Gender, Class, and Social Desirability." *Gender & Society* 12(2):188–218.

Ray, John J. 1984. "The Reliability of Short Social Desirability Scales." *The Journal of Social Psychology* 123(1):133–134.

Ripley, Elizabeth, Francis Macrina, Monika Markowitz and Chris Gennings. 2010. "Who's Doing the Math? Are We Really Compensating Research Participants?" *Journal of Empirical Research on Human Research Ethics* 5(3):57–65.

Rosenfeld, Bryn, Kosuke Imai and Jacob N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* 60(3):783–802.

Sârbescu, Paul, Iuliana Costea and Silvia Rusu. 2012. "Psychometric Properties of the Marlowe-Crowne Social Desirability Scale in a Romanian Sample." *Procedia - Social and Behavioral Sciences* 33:707–711.

Schoderbek, Peter P. and Satish P. Deshpande. 1996. "Impression Management, Overclaiming, and Perceived Unethical Conduct: The Role of Male and Female Managers." *Journal of Business Ethics* 15(4):409–414.

Sexton, Jared. 2008. *Amalgamation Schemes: Antiblackness and the Critique of Multiracialism*. Minneapolis: University of Minnesota Press.

Shulevitz, Judith. 2016. "In College and Hiding from Scary Ideas." *New York Times*. March 21, 2016 .

Slothuus, Rune and Claes H. De Vreese. 2010. "Political Parties, Motivated Reasoning, and Issue Framing Effects." *The Journal of Politics* 72(3):630–645.

Stoudt, Brett G. 2007. "The Co-Construction of Knowledge in "Safe Spaces": Reflecting on Politics and Power in Participatory Action Research." *Children Youth and Environments* 17(2):280–297.

Strahan, Robert and Kathleen Carrese Gerbasi. 1972. "Short, Homogeneous Versions of the Marlow-Crowne Social Desirability Scale." *Journal of Clinical Psychology* 28(2):191–193.

31

Theriault, Stephen W. and Diane Holmberg. 1998. "The New Old-Fashioned Girl: Effects of Gender and Social Desirability on Reported Gender-Role Ideology." *Sex Roles* 39(1-2):97–112.

Wharton, Amy S. 2009. "The Sociology of Emotional Labor." *Annual Rreview of Sociology* 35:147–165.