**Safe Here, but Unsafe There? Institutional Signals of Identity Safety Also Signal Prejudice Elsewhere**

by

Izzy Gainsburg

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2020

Doctoral Committee:

Assistant Professor Allison Earl, Chair
Associate Professor Sonya Dal Cin
Professor Ethan Kross
Professor Denise Sekaquaptewa

Izzy Gainsburg

izzyg@umich.edu

ORCID iD:  0000-0003-4363-0494

## Dedication

To my mother, Phyllis, for her motherly wisdom; to my father, Danny, for his relentless curiosity; to my brother, Jesse, for setting the best example; and to my partner, Briana, for her unwavering love and support.

**Acknowledgements**

Discovering one's place in the world, getting into graduate school, and actually succeeding once there—none of it is easy. It takes tremendous luck and support from others. Because writing an acknowledgments section about luck would be dry, I'd like to elaborate on how grateful I am for the social support that guided me to my doctorate.

Mom and dad—you instilled in me a moral code and curiosity that has infused my work. But more importantly, you have been there for me when things were most difficult, from thinking through career decisions to helping me cope with setbacks. Jesse—your principled approach to life's most important problems has made me a better scientist. But what I'm most grateful for is being able to chat with you every day about basketball, biking, food, ethics, or whatever else is top of mind (but let's face it, mainly basketball).

The first person to set me on the psychology path was my AP Psychology teacher, Bobby Asher. Bobby—I enjoyed your class more than any other through high school. At Tufts, my advisor and teacher Sam Sommers had a similar effect on me. Sam—I'm not sure there is anyone on earth as good as you at translating social psychology for a college audience, and it's been a pleasure connecting more since I've continued on at Michigan. China Gate forever! I would also like to thank Nalini Ambady, Kristin Pauker, Max Weisbuch, and Anne Krendl, who gave me the license to design and run my own experiments, which was necessary experience for continuing forward in psychology.

Graduating college was scary for many reasons, but mostly because I didn't have a job. I eventually hit the jackpot, though, by getting the opportunity to work as a research specialist

under Danny Oppenheimer, Eldar Shafir, and Betsy Levy Paluck. Eldar—you opened my eyes to how psychology can inform policy around the world's biggest problems. Betsy—you have taught me to not set limits for what psychological research and intervention can look like. And to those that reviewed my graduate school application materials—Rebecca Littman, Kyle Keller, Hana Shepherd, Robin Gomila, Abigail Sussman, among others—here's another thank you some seven years later.

Of course, the majority of "social support credit," as far as earning my doctorate is concerned, goes to the people I connected with during graduate school. The University of Michigan psychology department has an amazing staff and faculty, a warm culture, and just about every resource a scholar could hope for. Most of all, I want to thank my advisors. You guys have set quite the example, from running amazing lab meetings to hosting memorable (and vegan-friendly!) barbeques, potlucks, and happy hours.

To my most recent mentor, Julia Lee Cunningham--it's been an incredible experience working with you over the past year. As was the case with getting my job at Princeton or getting into University of Michigan, I feel like once again, I've hit the jackpot by getting the chance to work together. To my first Michigan mentor, Denise Sekaquaptewa--you've made my work more impactful by setting an example of scientific rigor and pushing me to consider the real-world impact of my work. I'll never forget the moment you broke out the Slivovitz! To Ethan Kross, who has also supported me since day one--you've taught me how to think precisely, but also, how to think big. Talking through ideas with you—on walks, in your office, or in the classroom—is usually fruitful and always fun. And finally, to my dissertation chair, Allison Earl—this roller coaster of a project would not have finished its loop without you keeping it on track. But we've managed to do just that, and it's thanks to your patience and pushing me to

think on a deeper, more theoretical level. Call it corny, but there is something special about Veronica and I graduating at essentially the same time that you earn your tenure promotion.

I'd be remiss not to mention others who directly contributed to the dissertation. Sonya Dal Cin—you were my first glimpse of Wolverine brilliance outside the psychology department and went above and beyond in refining this project. Thank you to the many research assistants and lab managers who have helped me with this research, with a special thank you to Joseph Calabrisotto, Anna Perrier, Samuel Emblidge, Kalei Glozier, Steven Cioffi, and Amena Khan. The fingerprints of your creative minds and work ethic are all over this dissertation. And of course, thank you to my lab mates in the HAILab, SPRIG Lab, and ESC lab, who have helped me think through this project countless times over.

Naturally, many of those lab mates were my friends. Graduate school is a great time to make new friends, but it only works if you are surrounded by awesome people. From the moment I arrived as a naïve first-year, friends were there to make sure I never got lost. To the older students who showed me the Big House ropes—Joe Bayer, Dave Hauser, Josh Wondra, Walter Sowden, Darwin Guevarra, Lauren Reed, Sarah Huff, Steve Tompson, and Ben Blankenship—it feels like just yesterday we were cheering on Devin Gardner. Luckily, Michigan people in the younger cohorts think I'm alright—Julia Rios, Wilson Merrell, Nicholas Michalak, and Todd Chan, it's been awesome getting to know you guys over the years. To my office mates--Peter Felsman, Koji Takahashi, Neil Lewis, and Martha Berg—you all have been a sounding board for every research thought I've ever had, and have doubled as a roommate/pick-and-roll partner, recruitment guest/statistical consultant, and collaborator/mentor. To my cohort—Ariana Orvell, Veronica Derricks, and Meg Seymour—we did it! I still remember going to Mash to celebrate finishing our first year. You guys have been such an important part of me surviving this

journey. To my outside-of-psychology friends—Jeff, Kate, Eric, Sarah, Kassia, Michael, Danny, Michelle, Will, and everyone else—thank you for tolerating a nerdiness I can't seem to turn off. And finally, to Hakeem Jefferson and Michael Hall. Mike—I still remember the warm fuzzies I had when you hosted me during recruitment weekend. And thank the lord Yeezus, you introduced me to Hakeem that same weekend. What a goddamn run we had, hanging at the Grotto, dancing the nights away, and kicking back at the Betsie River Lodge.

Finally, I want to acknowledge Briana Green and the love and support she has given me throughout graduate school. Briana—you edited my 619 paper within a month of knowing me, and I've leaned on you for your editing skills more times than I'd care to admit. But you don't mind, which makes sense, because you have been there for me in every other way throughout graduate school, too. You've helped me cope with study failures, anxieties about my research identity, and uncertainties around career decisions—in psychology terms, you've been the perfect balance of instrumental and emotional social support. More than that, though, you lead by example. Your ambition, work ethic, and well-roundedness inspire me every day (especially now that we work from home together due to the coronavirus pandemic). I really do see this as a shared accomplishment. As much as anything at school, my time here has been defined by the home we've created with our favorite kitty friend, Willow. Lucky for us, we'll be here another year!

## Table of Contents

## List of Tables

TABLE

## List of Figures

**Abstract**

Across four studies ($N$ = 3049, 69.7% White, 45.5% Male, 85.2% Heterosexual), participants who read about institutions with signals to identity safety (compared to signals unrelated to identity safety) showed increased expectations of prejudice in other environments. These increases in prejudice expectations mediated increased perceptions of prejudice (Study 1), motivations to combat prejudice (Study 2 and 3), and support for movements and policies that address diversity, equity, and inclusion (Study 3 and 4). Moreover, these increased prejudice expectations were not moderated by group membership. Evidence suggested that signals to identity safety increased prejudice expectations in non-signaled spaces because they are perceived as a response to problem in the broader environment, leading people to infer the existence of a problem. In addition, across studies, signals to identity safety were interpreted differently than institutional signals of other kinds of social problems—intuitional signals of environmental friendliness, for instance, did not increase perceptions of environmentally harmful behavior in other environments. Collectively, the current studies broaden the scope of prior work on identity safety signals by showing that identity safety signals transfer across location and that prejudice expectations can lead to both negative outcomes (e.g., perceptions of prejudice) and positive outcomes (such as support for movements/policies that promote diversity, equity and inclusion).

*Keywords:* prejudice, identity threat, expectations, environmental cues

## Chapter 1

Institutions and organizations often take measures to create safe and inclusive environments, especially for people belonging to underrepresented or marginalized groups. These efforts range from commitments to diversity to anti-prejudice policies to designated spaces or events for specific social identity groups. Although these measures are intended to improve organizational climate (e.g., by reducing prejudice), they are also intended to make organizations *feel* safe. In other words, these efforts function as signals for what kind of social climate one can expect in the institution (Chaney & Sanchez, 2018; Chaney, Sanchez, & Remedios, 2016; Cundiff, Ryuk, & Cech, 2018; Murphy & Destin, 2016; Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby, 2008).

Prior research has focused on how signals of identity safety affect people's feelings and perceptions of identity threat and safety inside a signaled space (e.g., Chaney et al., 2016; Cundiff et al., 2018). Furthermore, this work has examined whether these identity safety signals transfer across identities. Chaney and colleagues (2016), for instance, showed that White women expected to feel included more in organizations that had training programs targeted at fostering the success of racial minorities. Cundiff and colleagues (2018), on the other hand, showed that men were more concerned with being treated negatively on account of their gender when a company had women-focused diversity initiatives. Thus, it is clear that organizational cues to identity safety targeted at one group can transfer to impact members of social identity groups.

What remains unclear, however, is how organizational cues to identity safety transfer across other kinds of dimensions, such as time or space. To address this gap, the present research

explores whether organizational cues to identity safety transfer across physical locations. In other words, it is unclear how institutional signals of identity safety impact people's expectations of identity threat in other locations, i.e., in areas *outside* the space signaled to be safe. For instance, encountering a university multicultural center might increase or decrease students' expectations of prejudice in other places (e.g., the rest of campus, the United States). Because people's day-to-day lives often occur outside explicitly signaled spaces, it is important to understand whether signals of identity safety also affect people's expectations of identity threat in these other environments, given that threat expectations can have important downstream consequences that are both negative (e.g., effects on stress and anxiety; Major, Mendes, & Dovidio, 2013) and positive (motivations to reduce prejudice; Mallett, Huntsinger, Sinclair, & Swim, 2008; Paluck, 2011).

Thus, the present research tests how institutional signals of identity safety affect expectations of identity threats in other environments, addressing an important theoretical and practical gap in research on organizational efforts to create safe and inclusive social climates.

**Institutional Signals of Identity Safety**

Social identity threats are concerns about being unfairly judged based on the social identity groups to which one belongs (Steele, Spencer, & Aronson, 2002). A number of organizational cues, such as underrepresentation of one's social identity group (Cohen & Swim, 1995; Inzlicht & Ben-Zeev, 2000, 2003; Niemann & Dovidio, 1998; Sekaquaptewa & Thompson, 2002, 2003) or physical features of the environment (e.g., posters on a wall) that signal belonging for some groups more than others (Cheryan, Plaut, Davies, & Steele, 2009; Murphy, Steele, & Gross, 2007), can lead people to expect or experience identity threat (for a review, see

Emerson & Murphy, 2014). Because many organizations aim to reduce identity threat, they signal that they are safe and inclusive environments for all social identity groups.

Some signals of safety include organizational programming, such as diversity trainings or poster campaigns. A prominent example of this was Starbucks closing all its stores to implement a diversity training in response to an employee who denied a black patron's request to use the bathroom (Donnelly, 2018). Another example is University of Michigan's "Expect Respect" campaign, which uses posters to promote tolerance and discourage prejudice (Shaikh, 2018). Empirical research on these kinds of organizational initiatives has produced mixed effects. For instance, diversity messages can improve trust and comfort in an organization among minority groups (Purdie-Vaughns et al., 2008), but can alienate majority groups (Brief et al., 2005; Dover, Major, & Kaiser, 2016; Kalev, Dobbin, & Kelly, 2006) and lead majority groups to believe that an organization is fair, leading them to discount prejudice faced by minority group members (Kaiser et al., 2013).

Another type of identity safety signal includes organizational policies, such as those that punish prejudice or promote equality in hiring and promotion. These signals, too, can have mixed effects on identity threat expectations. For instance, organizational awards honoring the success of people from underrepresented groups can increase identity safety for those group members (Chaney et al., 2016). On the other hand, affirmative action policies and equal employment opportunity statements can elicit concerns among people from underrepresented groups about whether they were hired on account of their own merit (i.e., attributional ambiguity; Leibbrandt & List, 2018; Major, Feinstein, & Crocker, 1994). Taken together, efforts to promote identity safety can have many consequences, motivating the present research around how they affect threat expectations in other, non-signaled environments.

One type of institutional signal that has not received attention in the social psychology literature is *safe spaces*—designated areas where people can feel comfortable in their social identities and safe from prejudice. Safe spaces have existed since at least the 1960's, at which time they served as protective and communal spaces for people in the LGBTQ community (Hanhardt, 2013; Kenney, 2001). Since then, safe spaces have grown more popular and varied—they exist in both schools (Holley & Steiner, 2005) and workplaces (Hill, 2009) and for both specific groups and for non-specific populations (Arao & Clemens, 2013). One feature of safe spaces is that they are localized, making them ideal for testing how identity safety signals affect people's expectations inside and outside signaled spaces.

**Why Prejudice Expectations Are Important**

As mentioned earlier, research on organizational signals of identity safety has been limited to their effects with respect to the signaled space. These signals, however, may also inform people's expectations of identity threat in the broader environment. One widely studied identity threat is expectations of prejudice (Inzlicht, Kaiser, & Major, 2008; Shelton, Richeson, & Salvatore, 2005). Prejudice expectations can have important downstream consequences for people of both privileged and marginalized groups. When marginalized group members expect prejudice, they are more vigilant for prejudice (Kaiser, Vick, & Major, 2006), are more likely to interpret ambiguous behaviors as prejudiced (Operario & Fiske, 2001), experience higher rates of stress and anxiety (Major et al., 2013), and show adverse health outcomes (Sawyer, Major, Casad, Townsend, & Mendes, 2012). Prejudice expectations also have more positive effects, such as increased intentions to reduce prejudice and stereotyping (Mallett et al., 2008; Paluck, 2011). This is consistent with theories of critical consciousness (Kumagai & Lypson, 2009) and multicultural education (Abrams & Gibson, 2007), in which understanding and awareness of

prejudice foster motivation to combat prejudice and increase egalitarianism (Case, 2007; Stewart, Latu, Branscombe, Phillips, & Denney, 2012). The present research focuses on two consequences of prejudice expectations—perceptions of prejudice in ambiguous behaviors and egalitarian attitudes and motivations—both of which might follow for people from marginalized and privileged groups.

**Effect of Institutional Signals of Identity Safety on Prejudice Expectations**

Because institutional identity safety signals are often intended to reduce expectations of prejudice inside the signaled space, they should reduce prejudice expectations inside that location. The primary goal of the present research, however, is to explore how institutional signals to identity safety influence prejudice expectations *outside* of the signaled space. These signals could increase *or* decrease prejudice expectations outside of the signaled space, and the present research tests these competing hypotheses.

Signals of identity safety might reduce prejudice expectations in outside spaces, for instance, if their effect of reducing prejudice inside the signaled space transfers to adjacent environments (i.e., the *transferability hypothesis*). Supporting the *transferability hypothesis*, the benefits of identity safety signals for one demographic group can transfer to other groups; for instance, women reading about companies with racial diversity trainings (compared to a control group) expect these companies to be fairer for women, too (Chaney et al., 2016). Similarly, the effect of an identity safety signal might transfer across space: people may infer identity safety and reduce prejudice expectations for spaces adjacent to signaled locations.

Whether the effects of identity safety signals transfer to outside environments may be modulated by whether the signaled spaces are typical of the outside environment. This hypothesis is related to cognitive assimilation (Bless & Schwarz, 2010). For instance,

participants who first considered two "typically favorable" television shows subsequently evaluated television shows in general as more favorable, compared to participants in a control condition (Bless & Wänke, 2000). Similarly, if people perceived signaled spaces as safe and typical of the broader environment, they may expect less prejudice in the broader environment. However, if people perceive the signaled space as atypical of the broader environment, people may perceive the broader environment as unsafe by comparison and expect *more* prejudice in the broader environment (i.e., *the atypicality hypothesis*). This is related to cognitive contrast (Bless & Schwarz, 2010). In the previously discussed study (Bless & Wänke, 2000), participants who first considered two "atypically favorable" television shows subsequently evaluated television shows in general as less favorable, compared to participants in a control condition. Thus, if people perceive signaled spaces as safe but atypical of the broader environment, they may expect more prejudice in the broader environment.

Finally, institutional identity safety signals might increase prejudice expectations in outside environments if they are perceived as a response to a problem in the broader environment (i.e., *the response to a problem hypothesis*). This is similar to the *atypicality hypothesis*—in both cases, broader environments are seen as unsafe in comparison to a signaled space. The *response to a problem hypothesis* involves a key difference, however. For the *atypicality hypothesis,* increased prejudice expectations in outside spaces occur through mere contrast; for the *response to a problem hypothesis,* increased prejudice expectations in outside spaces happen through logical inference. These mechanisms predict different outcomes under certain conditions. Specifically, the *atypicality hypothesis* would not predict identity safety signals to increase prejudice expectations in the broader environment if the signaled spaces are in broader spaces with other identity safety signals, and thus typical of the broader environment. The *response to a*

*problem hypothesis*, however, predicts that safety signals (even if abundant and typical of the environment) are evidence of a problem, thus increasing prejudice expectations in the broader environment.

**Moderation by group membership.** Institutional identity safety signals are often intended to reduce identity threats for marginalized group members. Furthermore, the effects of environmental cues on threat perception often differ across groups (Emerson & Murphy, 2014). Thus, we test whether the effects of identity safety signals on prejudice expectations differ for historically marginalized (vs. privileged) groups. Identity safety signals may have a weaker effect on prejudice expectations in outside spaces for marginalized (vs. privileged) groups, given that prejudice concerns are more chronically active for members of marginalized groups (Mendoza-Denton, Downey, Purdie, Davis, A., & Pietrzak, 2002), overwhelming the effects of identity safety signals. On the other hand, identity safety signals may have stronger effects for marginalized (vs. privileged) groups; for instance, if calibrating prejudice expectations is more important for members of marginalized (vs. privileged) groups, people from marginalized groups may make more use of the cue to inform their expectations of threat in other environments.

## The Present Research

The first goal of the present research is to test how exposure to institutional signals of identity safety influence prejudice expectations inside and outside the signaled space. We tested this with two signals—university safe spaces (Study 1, 2, and 4) and company commitments to diversity (Study 3). In addition, we conducted exploratory analyses to examine if the effects of identity safety signals differ across groups.

The second goal of the present research is to explore the process through which identity safety signals affect prejudice expectations for non-signaled spaces. In particular, Studies 2 and 4

were designed to examine the *atypicality* and *response to a problem* hypothesis, respectfully. Across studies, we test whether these mechanisms are unique to institutional identity safety signals, or if they apply to other environmental signals as well (e.g., signals of cell phone use; signals of environmental behavior).

The third goal is to test whether prejudice expectations mediate various downstream consequences of exposure to identity safety signals, including perceptions of ambiguous behaviors as prejudiced (Study 1), egalitarian attitudes and motivations (Study 2, 3, and 4), and attitudes toward movements/policies that support diversity, equity, and inclusion (Study 3 and 4), and donations to an anti-prejudice organization (Study 4).

Collectively, the current studies are the first to test whether organizational safety cues transfer across location. Prior work has demonstrated that institutional signals of identity safety can transfer across groups (Chaney et al., 2016), but we advance this work by testing how such cues transfer across physical environments. In doing so, we test the direction in which these cues transfer (i.e., whether they result in increased or decreased prejudice expectations) and potential mechanisms underlying these effects (e.g., whether these signals are perceived as a response to a problem). Moreover, the current studies broaden the scope of prior work on the consequences of identity safety signals and prejudice expectations by exploring how those expectations can motivate both negative outcomes (e.g., perceptions of prejudice) and positive outcomes (such as support for movements/policies that promote diversity, equity and inclusion)..

## Study 1a, 1b, and 1c

We used an identical procedure for Studies 1a, 1b, and 1c. Participants were randomly assigned to read a vignette where they either encountered a safe space or “no cell phone” space

at a university. We used a university as the institution type because that is a common environment for safe spaces.

We used Amazon's Mechanical Turk (MTurk) for Study 1a to collect data quickly. We then replicated the findings with college undergraduates (Study 1b). Finally, we re-ran the study on MTurk immediately following the 2016 United States presidential election to test election-relevant exploratory hypotheses (Study 1c). These hypotheses are unrelated to the present research and are not further explored. Because results are similar for all three studies, we present analyses collapsed across dataset.

## Method

### Participants

We set out to recruit 400 people in each study. We chose 400 as a target sample size to achieve sufficient power (.80) to detect a small-medium effect size of $d = .30$ and to account for potential attrition and exclusions. In the two MTurk studies, additional participants provided usable data despite not finishing the survey. In the study with college undergraduates, data collection stopped at semester's end, leaving us slightly under our target sample size. Analyses were conducted following data collection. See Table 1 for demographics from Study 1 and Table 2 for more detailed demographics from all studies.

Table 1

*Study 1 Sample and Demographics*

| | Original *N* | Excluded | Final *N* | Female % | Non-White % | Non-heterosexual % | Minimum *d* |
|---|---|---|---|---|---|---|---|
| Study 1a | 434 | 36 | 398 | 50.3 | 25.6 | 13.6 | 0.28 |
| Study 1b | 336 | 48 | 288 | 53.8 | 40.3 | 5.9 | 0.33 |
| Study 1c | 429 | 38 | 391 | 46.3 | 22.8 | 16.4 | 0.28 |
| Total | 1199 | 122 | 1077 | 49.8 | 28.5 | 12.5 | 0.17 |

Exclusions were based on not consenting ($N = 11$), dropping out before answering any questions ($N = 22$), failing a manipulation-relevant attention check ($N = 87$), and being non-native English speakers ($N = 2$). Percentages are of Final *N,* not of valid respondents to demographic items. Minimum *d* is based on sensitivity power analysis in

GPower (Faul, Erdfelder, Lang, & Buchner, 2007) for an independent samples t-test, given Power = .80 and α = 0.05. All subsequent power analyses also used G*Power.

Table 2

*Demographic information from Studies 1-4*

| | | Study 1 | | Study 2 | | Study 3 | | Study 4 | |
|---|---|---|---|---|---|---|---|---|---|
| *Race* | | N | Percent | N | Percent | N | Percent | N | Percent |
| | White/Caucasian | 730 | 67.8% | 400 | 66.4% | 400 | 73.0% | 596 | 72.5% |
| | African American/Black | 71 | 6.6% | 52 | 8.6% | 47 | 8.6% | 76 | 9.2% |
| | Hispanic/Latino | 35 | 3.2% | 27 | 4.5% | 34 | 6.2% | 35 | 4.3% |
| | Native Hawaiian/Pacific Islander | 0 | 0.0% | 2 | 0.3% | 0 | 0.0% | 0 | 0.0% |
| | Native American/American Indian | 9 | 0.8% | 6 | 1.0% | 5 | 0.9% | 4 | 0.5% |
| | Asian/Asian American | 88 | 8.2% | 59 | 9.8% | 29 | 5.3% | 44 | 5.4% |
| | South Asian/South Asian American | 20 | 1.9% | 1 | 0.2% | 2 | 0.4% | 3 | 0.4% |
| | Middle Eastern American/Arab | 12 | 1.1% | 2 | 0.3% | 1 | 0.2% | 1 | 0.1% |
| | Multiracial | 67 | 6.2% | 43 | 7.1% | 23 | 4.2% | 51 | 6.2% |
| | Other | 5 | 0.5% | 1 | 0.2% | 1 | 0.2% | 2 | 0.2% |
| | Missing/No Response | 40 | 3.7% | 9 | 1.5% | 6 | 1.1% | 10 | 1.2% |
| *Gender* | | | | | | | | | |
| | Men | 492 | 45.7% | 248 | 41.2% | 261 | 47.6% | 385 | 46.8% |
| | Women | 536 | 49.8% | 341 | 56.6% | 281 | 51.3% | 422 | 51.3% |
| | Other | 12 | 1.1% | 2 | 0.3% | 3 | 0.5% | 9 | 1.1% |
| | Missing/No Response | 37 | 3.4% | 11 | 1.8% | 3 | 0.5% | 6 | 0.7% |
| *Sexual Orientation* | | | | | | | | | |
| | Heterosexual | 901 | 83.7% | 526 | 87.4% | 476 | 86.9% | 695 | 84.5% |
| | Non-heterosexual | 135 | 12.5% | 66 | 11.0% | 64 | 11.7% | 114 | 13.9% |
| | Missing/No Response | 41 | 3.8% | 10 | 1.7% | 8 | 1.5% | 13 | 1.6% |

**Procedure**

After informed consent, participants were told the study was about "people's thoughts about different kinds of signs and the types of spaces they represent." Next, participants were told that they would learn about different types of signs that they might answer questions about later in the study. Participants were then randomly assigned to either the safe space condition or the control condition. In both conditions, participants learned about four signs, each presented on a separate page. For each sign, participants were given an example picture and a description of what it meant.

In the safe space condition, participants saw “Cell Phone Ban” signs first, “Quiet Space” signs second, “Safe Space” signs third, and “Germ Free” signs fourth. The safe space sign was described as “these types of signs are indicative of spaces where everyone can feel comfortable about expressing their identity without fear of discrimination or attack,” which was a paraphrased definition from Google’s definitions (“Safe Space”, 2016). In the control condition, participants saw “Smoke Free” signs first, “Quiet Space” signs second, “Cell Phone Ban” signs third, and “Germ Free” signs fourth.

Note that in this design, participants in both conditions read about the “Cell Phone Ban” signs; those in the Safe Space condition read about the “Safe Space” sign (but not the “Smoke Free” sign), and those in the control condition read about the “Smoke Free” sign (but not the “Safe Space” sign). We opted for this design because it was consistent with our goal of preventing participants from knowing that our study was about Safe Spaces. In other words, this design allowed us to have a control condition where Safe Spaces were not at all salient for participants. We nonetheless had participants in the Safe Space condition read about the “Cell Phone Ban” sign in order for our filler questions about the prevalence of cell-phone use to feel natural—we did not want participants in this condition to know that the task was about Safe Spaces. Moreover, by keeping the Cell Phone sign (and questions about cell phone use) in both conditions, we were able to hold even more information constant across conditions. (this was not A limitation of this design is that it makes it difficult to interpret whether differences between the “Cell Phone Ban” and “Safe Space” signs for sign-relevant expectations are due to the type of signal or due to the fact that all participants saw one sign but not the other. We address this in Study 3 by randomly assigning participants to conditions where only read about an institutional identity safety signal or a control signal.

Participants then read a vignette about a college student named John waiting in his advisor's office for a meeting. In the safe space condition, John notices a number of things in the office as he waits, including a "Safe Space" sign; in the control condition the sign says "No Cell Phone Use." The vignettes were otherwise identical and are in the appendix. Participants next indicated which signs were mentioned in the vignette (an attention check) and rated the prevalence of cell phone use and prejudice in the advisor's office, on John's college campus, and in the United States.

Next, participants were told they were going to read 16 scenarios for use in a future survey about human behavior. The scenarios contained ambiguous prejudice, blatant prejudice, or benign behavior. For each scenario, participants reported whether it contained a microaggression, implicit prejudice, and explicit prejudice. Participants read definitions of these terms prior to reading the scenarios. A microaggression was defined as "a subtle but offensive comment or action directed at a minority or other nondominant group that is often unintentional or unconsciously reinforces a stereotype." Implicit bias was defined as "attitudes and beliefs about a person or group that operate at a level below conscious awareness and without intentional control." Explicit bias was defined as "attitudes and beliefs about a person or group that individuals are consciously aware of." Each scenario and its associated questions were presented on separate pages in a randomized order.

We chose 16 scenarios from a piloted set of 32 based on a number of considerations. First, we wanted scenarios that varied in severity including ambiguous prejudice (e.g., a physics professor overlooking a female student), blatant prejudice (e.g., an internet commenter berating a Black commenter on account of their race), and benign behavior (e.g., a cashier giving change to a customer). We included this variability so that participants would think critically about each

scenario and also to demand characteristics by including scenarios where they could report that there is no bias. Second, we chose scenarios with prejudice against different groups (anti-Black, anti-Asian, anti-woman, and anti-LGB). Third, we used Sue et al.'s (2007) classification of microaggressions to generate different ambiguous scenarios, including microinsults (e.g., assumption of criminality) and microinvalidations (e.g., denial of individual racism), and chose scenarios depicting distinct microaggressions. Finally, we favored scenarios rated as realistic by pilot participants. The final set included two blatant scenarios (one anti-Black, one anti-Woman), two benign scenarios (one Black false target, one woman false target), and 12 ambiguous scenarios, (3 anti-Black, 3 anti-Asian, 4 anti-woman, 2 anti-LGB). See appendix for the scenarios.

Finally, participants answered questions about how often they see safe spaces, their attitudes toward political correctness, their beliefs about safe spaces as coddling or protective, and demographic items.

**Measures**

Relevant measures are listed below. All other measures are in the appendix.

**Prevalence of prejudice in advisor's office.** Participants answered, "How prevalent do you think prejudice (e.g., racism, sexism, homophobia, etc.) is within John's advisor's office / on John's college campus / in the United States?" for each location, using a sliding scale from 0 (Not at all prevalent) to 100 (Extremely prevalent).

**Prevalence of cell phone use.** Participants answered, "How prevalent do you think cell phone use is within John's advisor's office / on John's college campus / in the United States?" for each location, using a sliding scale from 0 (Not at all prevalent) to 100 (Extremely prevalent).

**Microaggressions in scenarios.** Participants answered, "From your point of view, to what extent would you characterize _____ as a microaggression?" from 1 (Not at all a microaggression) to 7 (Definitely a microaggression), with the scenario's relevant action in the blank.

**Implicit bias in scenarios**. Participants answered, "In your opinion how likely was it that _____ was motivated by implicit (unconscious) bias?" from 1 (Extremely unlikely) to 7 (Extremely likely), with the scenario's relevant action in the blank.

**Explicit bias in scenarios**. Participants answered, "In your opinion how likely was it that _____ was motivated by explicit (conscious) bias?" from 1 (Extremely unlikely) to 7 (Extremely likely)," with the scenario's relevant action in the blank.

## Results

We first tested if condition affected expectations of prejudice and cell phone use in the signaled space (the advisor's office) and the outside, non-signaled spaces (the college campus, and the United States; see Table 3 for descriptive and inferential statistics). We then tested direct and indirect effects on perceptions of prejudice in ambiguous behaviors, as mediated by prejudice expectations in non-signaled spaces (i.e., on campus and in the United States).

We also tested whether group membership (white vs. non-white, men vs. women, and heterosexual vs. non-heterosexual) moderated the effect of condition on prejudice expectations. The present research provided no consistent evidence for moderation by group membership despite our large sample. These analyses are in the appendix and are not explored in subsequent studies in the main text; we revisit this issue in the general discussion.

### Prejudice Expectations

Participants in the safe space (vs. control) condition expected no more or less prejudice in the signaled space (the advisor's office). However, participants in the safe space (vs. control) expected more prejudice in the non-signaled spaces (on campus and in the United States).

**Cell Phone Use Expectations**

Participants in the control (vs. safe space) condition expected less cell phone use in the signaled space (the advisor's office), marginally less cell phone use on campus, and no more or less cell phone use in the United States.

Table 3

*Expectations of cell phone use and prejudice from Study 1*

| | Control *M,* 95% CI | Safe Space *M*, 95% CI | *t* | *df* | *p* | *d* |
|---|---|---|---|---|---|---|
| Prejudice: Office | 16.70 [14.88, 18.52] | 17.49 [15.66, 19.32] | .60 | 1075 | .55 | .04 |
| Prejudice: Campus | 32.68 [30.69, 34.67] | 42.09 [40.09, 44.10] | 6.53 | 1075 | < .001 | .40 |
| Prejudice: US | 55.04 [53.06, 57.02] | 60.53 [58.54, 62.52] | 3.84 | 1075 | < .001 | .23 |
| Cell Use: Office | 34.32 [31.86, 36.78] | 57.97 [55.49, 60.45] | 13.28 | 1075 | < .001 | .81 |
| Cell Use: Campus | 86.11 [84.33, 87.89] | 88.47 [86.68, 90.26] | 1.83 | 1075 | .07 | .11 |
| Cell Use: US | 86.92 [85.22, 88.62] | 88.34 [86.64, 90.05] | 1.16 | 1075 | .25 | .07 |

**Perceptions of prejudice in ambiguous behaviors**

We predicted that participants in the safe space (vs. control) condition would perceive ambiguous behaviors as more prejudiced, and that this would be explained by elevated prejudice expectations in outside, non-signaled spaces. We tested this mediation model using PROCESS Model 4 (Hayes, 2017), which tests whether a predictor variable (i.e., condition: safe space vs. control) affects outcome variables (i.e., ratings of microaggressions, explicit bias, and implicit bias) via changes in a mediating variable (i.e., prejudice expectations in outside, non-signaled spaces). Specifically, the mediator was an average of participants' prejudice expectations on campus and in the United States (Cronbach's $\alpha = .74$). We used this as a mediator because it

broadly captured the outcome of interest in the present research (prejudice expectation in outside, non-signaled locations). See Figure 1 for Study 1 mediation analyses.

**Ratings of microaggressions.** Participants did not identify microaggressions any more or less in the safe space (vs. control) condition (i.e., the total effects were not statistically significant). However, there was a significant indirect effect: the safe space (vs. control) condition increased prejudice expectations in outside spaces, mediating increased ratings of microaggressions.

**Ratings of explicit bias.** Participants did not identify explicit bias any more or less in the safe space (vs. control) condition (i.e., the total effects were not statistically significant). However, there was a significant indirect effect: the safe space (vs. control) condition increased prejudice expectations in outside spaces, mediating increased ratings of explicit bias.

**Ratings of implicit bias.** Participants did not identify implicit bias any more or less in the safe space (vs. control) condition (i.e., the total effects were not statistically significant). However, there was a significant indirect effect: the safe space (vs. control) condition increased prejudice expectations in outside spaces, mediating increased ratings of implicit bias.

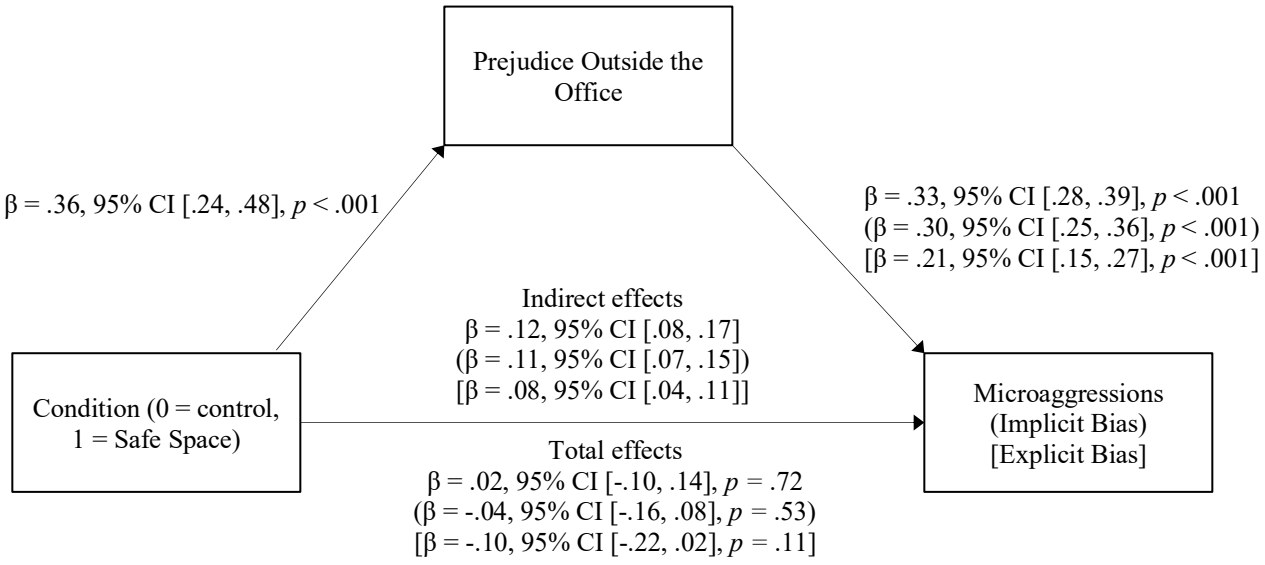


**Fig 1.** Effect of Safe Space cue on ratings of ambiguous behaviors as microaggressions, implicit bias, and explicit bias, as mediated by prejudice expectations outside the office (Study 1).

## Discussion

The first goal of the present research was to test how institutional identity safety signals affect prejudice expectations. Participants who read about a safe space (vs. no cell phone) sign in the office of a college advisor expected no more or less prejudice in the office, but expected more prejudice on campus and in the United States. We predicted that the safe space sign might reduce prejudice expectations inside the office, given that such expectations would be sign-consistent. In subsequent studies, we test whether this null finding replicates.

The "No Cell Phone" sign affected expectations of cell phone use differently from how the safe space sign affected prejudice expectations. The "No Cell Phone" sign strongly reduced expectations of cell phone use inside the advisor's office; this effect weakly extended to campus (although this effect was only marginal) and had little to no effect on estimates of cell phone in the United States. Put another way, the safe space sign did not influence expectations of sign-inconsistent behavior inside the signaled space, but increased expectations of sign-inconsistent behavior outside the signaled space; the "No Cell Phone" sign, on the other hand, reduced expectations of sign-inconsistent behavior inside the space and had little to no effect on expectations of sign-inconsistent behavior outside the signaled space. In addition, these findings were not moderated by group membership.

Study 1 also tested whether the safe space condition affected perceptions of prejudice in ambiguous behaviors. Reading about the safe space indirectly (but not directly) increased perceptions of prejudice in ambiguous behaviors via increased prejudice expectations in outside spaces. Although perceptions of prejudice can be threatening or stressful, prejudice expectations

may also have a positive effect by increasing motivations to combat prejudice, which we test in Study 2.

## Study 2

In Study 1, participants reading about a safe space (vs. a no cell phone) sign in an advisor's office expected more prejudice in spaces outside the office. One possible mechanism for this is the aforementioned *atypicality hypothesis.* The safe space sign could have made the advisor's office seem unusual and atypical of non-signaled locations, leading non-signaled locations to seem unsafe by comparison. If so, making the advisor's office typical of campus should mitigate increased prejudice expectations on campus. Thus, we adapted the Study 1 vignette in Study 2 such that the advisor's safe space sign is typical of campus.

Study 2 introduced two additional changes. First, the control condition used an "Eco-Friendly" sign, rather than a "No Cell Phone" sign, to give both the control and experimental condition positive language (vs. the prohibitory "No Cell Phone" sign) and to test whether the effects of the control condition generalized to other signs. Second, we measured egalitarian attitudes and intentions to confront prejudice as downstream consequences of prejudice expectations, rather than perceptions of prejudice, to demonstrate novel and positive effects flowing from identity safety signals and prejudice expectations.

### Method

#### Participants

The smallest significant main effect of condition in Study 1 was $d = .23$. We needed 596 participants to achieve 80% power of detecting an equivalent effect size. We oversampled this number to account for potential exclusions and recruited 624 individuals through MTurk. An additional 39 participants provided usable data despite not finishing the survey. Of these 663

participants, 61 were excluded (6 people did not consent, 8 dropped out before answering any questions, and 47 failed a manipulation-relevant attention check), leaving 602 participants in the final sample. This provided 80% power to detect an effect size of $d = .23$ for an independent samples t-test. Analyses were conducted following data collection.

**Procedure**

The Study 2 procedure was identical to Study 1 except for three differences. First an "Eco-Friendly" sign replaced the "No Cell Phone" sign in the sign definitions (from both conditions) and in the vignette in the control condition. Second, we made the advisor's office typical of campus by adding the sentence: "As he sees the sign, he realizes that he saw the same sign at the gym, the cafeteria, and in the offices of the other professors he's met with this semester." Third, participants reported their egalitarian attitudes and intentions to confront prejudice instead of reading the ambiguous prejudice scenarios.

**Measures**

Measures were identical to Study 1 with two exceptions. First, participants estimated prevalence of "environmentally-friendly behavior," instead of cell phone use. Second, participants reported egalitarian attitudes and intentions to confront prejudice using items created for the present research (described below).

**Intentions to confront prejudice.** Participants responded to three statements about their intentions to confront discrimination from 1 (Strongly Disagree) to 7 (Strongly Agree), e.g., "If I witnessed an act of discrimination, I would report it, if possible." These items were collapsed into a single measure (Cronbach's $\alpha = .87$).

**Egalitarian attitudes.** Participants responded to six statements about their attitudes toward diversity and egalitarianism from 1 (Strongly Disagree) to 7 (Strongly Agree), e.g.

"Being inclusive of people from different groups is important to me." These items were collapsed into a single measure (Cronbach's $\alpha = .89$).

## Results

We first test if condition affected expectations of prejudice and environmentally-friendly behavior in the advisor's office, the college campus, and the United States (see Table 4 for descriptive and inferential statistics; see appendix for tests of moderation by group membership). Next, we test direct and indirect effects on intentions to confront prejudice and egalitarian attitudes, as mediated by prejudice expectations in outside spaces (i.e., on campus and in the United States, the same mediator from Study 1; see Figure 2 for Study 2 mediation analyses).

### Prejudice Expectations

Replicating Study 1, participants in the safe space (vs. control) condition expected no more or less prejudice in the signaled space (the advisor's office) and expected more prejudice in the non-signaled spaces (on campus and in the United States).

### Environmentally-friendly Behavior

Replicating Study 1, participants in the control (vs. safe space) condition expected more sign-compliant behavior (i.e., more environmentally-friendly behavior) inside signaled space (the advisor's office). Outside the signaled space, participants in the control (vs. safe space) expected more environmentally-friendly behavior on campus and no more or less environmentally-friendly behavior in the United States.

Table 4

*Expectations of environmentally-friendly behavior and prejudice from Study 2*

| | Control *M,* 95% CI | Safe Space *M*, 95% CI | *t* | *df* | *p* | *d* |
|---|---|---|---|---|---|---|
| Prejudice: Office | 19.39 [16.82, 21.97] | 18.22 [15.61, 20.83] | .63 | 600 | .53 | .05 |
| Prejudice: Campus | 31.94 [29.21, 34.68] | 40.35 [37.59, 43.12] | 4.25 | 600 | < .001 | .35 |
| Prejudice: US | 52.45 [49.67, 55.22] | 60.77 [57.96, 63.59] | 4.14 | 600 | < .001 | .33 |
| Eco-friendly: Office | 73.22 [70.44, 75.99] | 50.84 [48.03, 53.65] | 11.12 | 600 | < .001 | .91 |
| Eco-friendly: Campus | 67.68 [65.13, 70.24] | 52.38 [49.79, 54.98] | 8.24 | 600 | < .001 | .67 |
| Eco-friendly: US | 44.33 [42.22, 46.43] | 46.05 [43.92, 48.19] | 1.13 | 600 | .26 | .09 |

**Intentions to Confront Prejudice**

Participants reported marginally higher intentions to confront prejudice in the safe space (vs. control) condition. This effect was mediated by prejudice expectations in outside spaces, such that the safe space (vs. control) condition increased prejudice expectations in outside spaces (i.e., campus and the United States), mediating increased intentions to confront prejudice.

**Egalitarian Attitudes**

Participants' reported significantly higher egalitarian attitudes in the safe space (vs. control) condition. This effect was mediated by prejudice expectations in outside spaces, such that the safe space (vs. control) condition increased prejudice expectations in outside spaces, mediating increased egalitarian attitudes.

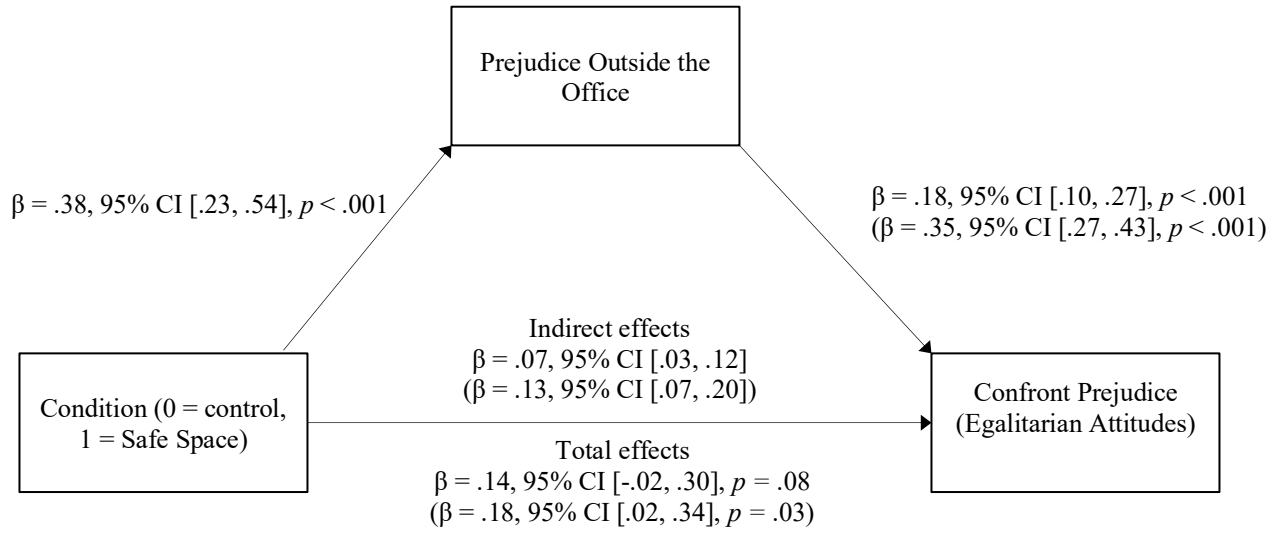



**Fig 2.** Effect of Safe Space cue on intentions to confront prejudice and egalitarian attitudes, as mediated by prejudice expectations outside the office (Study 2).

## Discussion

Study 2 tested whether making safe spaces typical of the broader environment mitigated increased prejudice expectations outside the safe space (i.e., the *atypicality hypothesis*). Instead, the pattern of results was consistent with Study 1: participants reading about an office labeled as a safe space (vs. an eco-friendly space), described to be typical of its broader environment, still expected more prejudice on campus (and in the United States), providing evidence against the *atypicality hypothesis*. Also consistent with Study 1, participants who read about a safe space (vs. a control) sign expected no more or less prejudice in the space with the cue.

The control sign—this time an eco-friendly sign—again worked differently than the safe space sign. The eco-friendly (vs. the safe space) sign increased expectations of environmentally-friendly behavior (i.e., a sign-consistent behavior) inside the signaled space. These increased expectations of sign-consistent behavior transferred to an adjacent space (i.e., participants also had increased expectations of environmentally-friendly behavior on campus, relative to the safe

space condition), but not to the United States. One reason that the safe space sign (but not the control signs) may have increased expectations of sign-inconsistent behavior is because the safe space sign is interpreted as a response to a problem in the broader environment (i.e., the aforementioned *response to a problem* hypothesis). This hypothesis is directly tested in Study 4.

Study 2 also tested whether encountering safe spaces increased intentions to confront prejudice and egalitarian attitudes. Participants in the safe space (vs. control) condition reported higher egalitarian attitudes and intentions to confront prejudice (mediated by increased prejudice expectations for outside spaces), demonstrating that identity safety signals can promote attitudes and behaviors that foster positive social climate and intergroup relations. However, it is unclear whether these effects are unique to safe spaces or whether other identity safety signals produce similar effects. Study 3 addresses this by testing how a different identity safety signal—a company's commitment to diversity—affects prejudice expectations and downstream consequences of those expectations.

## Study 3

Study 3 tested whether the findings from Study 1 and 2 generalize to a new signal and environment: a company's commitment to supporting diversity, equity, and inclusion (DEI). We chose a company as the context for three reasons. First, because the workplace is a focal context in previous research on identity safety and threat cues. Second, because it is relatable context for the Study 3 MTurk sample. Third, because using an entire company rather than a localized office addresses a potential floor effect in Studies 1 and 2. To clarify: it is possible that the low prejudice expectations in the signaled space (i.e., the advisor's office) in the control condition led to a floor effect, such that the identity safety signal (vs. the control signal) could not reduce prejudice expectations in the advisor's office. We speculated that by making the signaled space a

larger environment (i.e., a company), the identity safety signal might effectively reduce prejudice expectations inside the signaled space (relative to the control condition). More broadly, this change allows us to test how identity safety signals affect prejudice expectations in inside and outside spaces when the signaled space is larger. As far as the signal, we chose a DEI commitment because it fits the scope of identity safety signals in the present research (localized, intentional, institutional signals that do not target specific groups) and because they are common for companies.

## Method

### Participants

The smallest significant main effect of condition in Study 1 was $d = .23$. We needed 596 participants to achieve 80% power of detecting an equivalent effect size. We oversampled this number to account for potential exclusions and recruited 624 individuals through MTurk. An additional 44 participants provided usable data despite not finishing the survey. Of these 668 participants, 120 were excluded (3 people did not consent, 17 dropped out before answering any questions, and 100 failed a manipulation-relevant attention check), leaving 548 participants in the final sample. This provided 80% power to detect an effect size of $d = .24$ for an independent samples t-test.

### Procedure

Procedure was identical to the Study 2 except for two differences. First, participants were told the study was about "thoughts and feelings people might have when considering a new job" instead of being about types of signs. Second, the vignette was about the protagonist, John, waiting for a job interview at "Centium," a fictional marketing company. In the experimental

condition, John reads company materials as he waits, including their commitment to DEI; in the control condition, this is replaced with their commitment to sustainability.

### Measures

The measures were the same in Study 2 but adapted to the environments in Study 3. Participants judged the prevalence of environmentally-friendly behavior and prejudice in Centium, the city where Centium is located, other marketing companies, and the United States. In addition, participants reported their affirmative action attitudes using seven items (e.g., "Affirmative action is a good policy") from the Attitude Toward Affirmative Action Scale (Kravitz & Platania, 1993) from 1 (Strongly Disagree) to 5 (Strongly Agree). These items were collapsed into a single measure (Cronbach's $\alpha = .92$).

## Results

We first test if condition affected expectations of prejudice and environmentally-friendly behavior in each location (see Table 5 for descriptive and inferential statistics; see appendix for tests of moderation by group membership). Next, we test direct and indirect effects on intentions to confront prejudice, egalitarian attitudes, and affirmative action attitudes as mediated by prejudice expectations in outside spaces (a composite of prejudice expectations in the city where Centium exists, other marketing companies, and the United States; Cronbach's $\alpha = .87$; see Figure 3 for Study 3 mediation analyses).

### Prejudice Expectations

In Study 3, the "signaled space" was the company and the non-signaled spaces were other companies, the city where the company operated, and the United States. We predicted that participants in the DEI (vs. control) condition would expect no more or less prejudice inside the

signaled space, but would expect more prejudice outside the signaled space. Results generally confirmed these predictions.

**Environmentally-friendly Behavior**

Consistent with Study 2, participants reading about a space signaling eco-friendliness (vs. identity safety) expected more environmentally-friendly behavior inside the signaled space (the company), and this effect transferred to adjacent environments (with the effect weakening with increasingly distal environments).

Table 5

*Expectations of environmentally-friendly behavior and prejudice from Study 3*

| | Control *M,* 95% CI | DEI *M*, 95% CI | *t* | *df* | *p* | *d* |
|---|---|---|---|---|---|---|
| Prejudice: Company | 26.14 [23.29, 29.00] | 29.59 [26.82, 32.35] | 1.70 | 546 | .09 | .15 |
| Prejudice: City | 35.52 [32.69, 38.35] | 39.75 [37.01, 42.49] | 2.11 | 546 | .04 | .18 |
| Prejudice: Other Companies | 36.21 [33.47, 38.95] | 41.03 [38.38, 43.68] | 2.48 | 546 | .01 | .21 |
| Prejudice: US | 50.00 [47.01, 53.00] | 53.87 [50.97, 56.77] | 1.82 | 546 | .07 | .16 |
| Eco-friendly: Company | 74.28 [71.75, 76.81] | 53.84 [51.39, 56.29] | 11.40 | 546 | < .001 | .98 |
| Eco-friendly: City | 55.81 [53.41, 58.21] | 49.85 [47.53, 52.18] | 3.50 | 546 | .001 | .30 |
| Eco-friendly: Other Companies | 48.61 [46.17, 51.06] | 49.14 [46.77, 51.50] | .30 | 546 | .76 | .03 |
| Eco-friendly: US | 46.77 [44.50, 49.04] | 48.74 [46.55, 50.94] | .30 | 546 | .76 | .03 |

**Intentions to Confront Prejudice**

Participants' intentions to confront prejudice were no higher or lower in the DEI (vs. control) condition. However, there was a significant indirect effect, such that the DEI (vs. control) condition increased prejudice expectations in outside spaces, mediating increased intentions to confront prejudice.

**Egalitarian Attitudes**

Participants' egalitarian attitudes were no higher or lower in the DEI (vs. control) condition. However, there was a significant indirect effect, such that the DEI (vs. control)

condition increased prejudice expectations in outside spaces, mediating increased egalitarian attitudes.

**Affirmative Action Attitudes**

Participants' attitudes toward affirmative action were no higher or lower in the DEI (vs. control) condition. However, there was a significant indirect effect, such that the DEI (vs. control) condition increased prejudice expectations in outside spaces, mediating increased affirmative action attitudes.

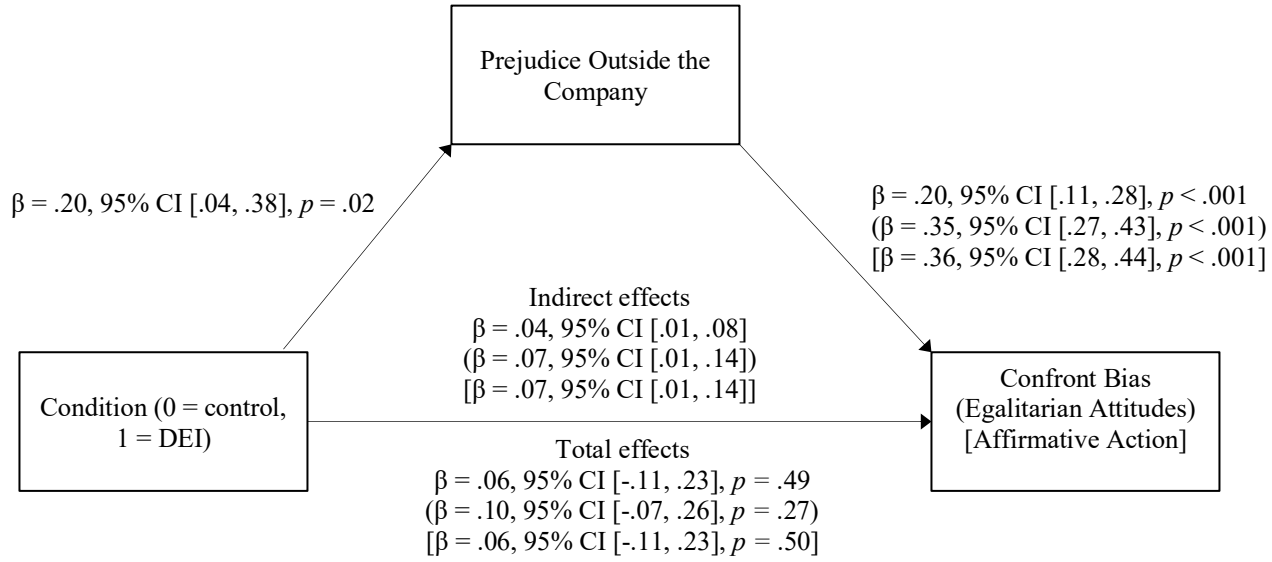



**Fig 3.** Effect of DEI cue on intentions to confront prejudice, egalitarian attitudes, and affirmative action attitudes, as mediated by prejudice expectations outside the office (Study 3).

## Discussion

Study 3 tested how a new signal (a company DEI statement) in a new context (the workplace) affected prejudice expectations. Consistent with Studies 1 and 2, participants in the safety signal (vs. control) condition expected more prejudice in areas outside the signaled space (i.e., the city where the company was located, other companies, and the United States); these

prejudice expectations in outside places again mediated increased intentions to confront prejudice and egalitarian attitudes, as well as affirmative action attitudes (new to Study 3). These findings are consistent with the *response to a problem hypothesis*, where the presence of identity safety signals indicates the presence of a problem, thereby increasing prejudice expectations, but Study 3 does not test this hypothesis directly. Study 4 addresses this gap by directly testing the *response to a problem hypothesis*.

We had speculated that one reason the safe space cue did not reduce prejudice expectations inside the office in Studies 1 and 2 was due to a floor effect. By making the signaled space larger, we thought that the identity safety cue might effectively reduce prejudice expectations inside the signaled space. However, despite the fact that the signaled space was a larger space in Study 3, we did not observe reductions in prejudice expectations inside the signaled space; if anything, the identity safety signal slightly increased prejudice expectation in the company. Although this effect was only marginal, it is consistent with the possibility that a safety signal may be perceived as a response to a problem inside the signaling institution, thus increasing prejudice expectations inside the signaling institution. This possibility would also be consistent with Study 2, where an identity safety signal increased prejudice expectations on campus despite the signal being prevalent on campus.

Finally, consistent with Study 2, a localized cue to eco-friendliness (vs. identity safety) increased expectations of environmentally-friendly behavior inside the signaled space, with this effect transferring to nearby environments (i.e., the city where the company operates) but not to more distal environments (i.e., other companies, the United States). To reiterate, this pattern is opposite to the one elicited by identity safety signals—in the previous three studies, identity safety signals increased expectations of sign-inconsistent behavior outside the signaled spaces

(but have little to no effect for the signaled space itself), whereas our control signals reduced expectations of sign-inconsistent behavior (Study 1) and increase expectations of sign-consistent behavior (Studies 2 and 3) inside the signaled space, with these effects transferring to adjacent non-signaled spaces (but not to more distant spaces).

## Study 4

Across the previous three studies, institutional signals to identity safety increased prejudice expectations in non-signaled environments, but had little to no effect on prejudice expectations inside the signaled environments. Other environmental cues produced opposite effects; in Studies 2 and 3, for instance, institutional signals of environmental-friendliness increased perceptions of environmentally-friendly behavior (i.e., a good thing) in the signaled environment and in adjacent (but not distant) non-signaled environments. It remains unclear, however, why institutional signals to identity safety increase prejudice expectations in non-signaled environments and why they produce a different pattern of expectations than other kinds of institutional signals.

We initially proposed two mechanisms that could explain why institutional identity safety signals increase prejudice expectations: *atypicality* and *response to a problem*. The *atypicality hypothesis* suggested that identity safety signals make signaled spaces seem unusual and atypical of non-signaled locations, leading non-signaled locations to seem unsafe by comparison. We ruled out *atypicality* in Study 2 because the participants in the safe space (vs. control) conditions expected more prejudice on campus, despite the safe space being typical of campus. The *response to a problem hypothesis* suggested that people perceive institutional identity safety signals as a response to a problem in the broader environment, thus inferring increased

prevalence of the problem in the broader environment from the signal. Study 4 tests this hypothesis in two ways.

First, we added a new condition that did not define safe spaces in terms of protection from discrimination, which was intended to reduce the inference that it was a response to a problem. The original description of safe spaces was "spaces where everyone can feel comfortable about expressing their identity without fear of discrimination or attack." The new condition described safe spaces as "spaces where everyone is welcomed and can feel comfortable about expressing their identity." We predicted that participants in the original (vs. new) safe space condition would rate the safe space as a response to a problem, mediating increased prejudice expectations in outside spaces.

Second, we directly measured participants' belief that the safe space sign is a response to a problem. We also directly measured participants' belief that the "Eco-friendly" sign was a response to a problem. We predicted the eco-friendly sign (vs. the safe space sign) would be seen as less of a response to a problem, helping explain why other signs have not increased expectations of problems in outside environments in previous studies.[1]

Study 4 extended the previous research in two other ways. First, participants indicated the groups they imagine receiving and enacting prejudice when reporting prejudice expectations. Although we suspected that all participants, regardless of group membership, were typically thinking of historically marginalized groups as receiving prejudice and historically privileged groups as enacting prejudice in prior studies, Study 4 directly tests this. Second, the present research introduced new attitudinal and behavioral measures of support for movements, policies,

[1] Previous studies asked about expectations of environmentally-friendly behavior, which did not decrease in outside spaces in the control (vs. safety signal) condition. Study 4 asks about "environmentally-harmful" behavior to frame it as a problem, like prejudice.

and organizations that support diversity, equity, and inclusion to show farther-reaching consequential outcomes of institutional identity safety signals.

## Method

### Participants

The smallest significant main effect of condition in Study 1 was $d = .23$. We needed 298 participants per condition (894 total) participants to achieve 80% power of detecting an equivalent effect size. We oversampled this number to account for potential exclusions and recruited 900 individuals through MTurk. An additional 37 participants provided usable data despite not finishing the survey. Of these 937 participants, 115 were excluded (6 people did not consent, 2 dropped out before answering any questions, and 107 failed a manipulation-relevant attention check), leaving 822 participants in the final sample.

### Procedure

The procedure in Study 4 was similar to Study 1, except for six things. First, a third of participants were randomly assigned to the new safe space condition. Second, participants answered questions addressing the *response to a problem hypothesis* before demographics. Third, the eco-friendly sign replaced the no cell phone sign in the sign definitions (in all conditions) and in the vignette in the control condition. Fourth, participants reported expectations of "environmentally-harmful" (instead of "environmentally-friendly") behavior. Fifth, participants did not evaluate ambiguous prejudice scenarios; instead, they reported attitudes toward well-known DEI policies and movements (e.g., the #MeToo movement) and how much they would be willing to donate to the Southern Poverty Law Center (an organization that works towards social justice, whose organizational mission was explained to participants). Sixth,

participants indicated the groups they thought were the targets and enactors of prejudice regarding their previously reported prejudice expectations.

## Measures

All new measures relevant to the present study are listed below.

**Sign is a response to a problem.** Participants rated their agreement with three items "The advisor put up his sign as a response to a problem [in his office] / [on campus] / [in the United States]" from 1 (Disagree) to 7 (Agree).

**Attitudes toward movements and policies.** Using a feeling thermometer (Brandt, Chambers, Crawford, Wetherell, & Reyna, 2015), participants rated their attitudes toward affirmative action, reducing the gender pay gap, the #MeToo movement, reparations for African-Americans, Black Lives Matter, laws preventing discrimination against LGBTQ people, and gender-neutral bathrooms. These items were collapsed into a single measure (Cronbach's $\alpha$ = .90).

**Donations toward Southern Poverty Law Center (SPLC).** As an incentive compatible measure, participants read two sentences about the SPLC and entered the amount of their payment that they were willing to donate (from $0.00 to $1.00).

**Groups that are targets of or enacting prejudice.** Participants separately indicated how much they were thinking about various groups (racial minorities, white people, women, men, LGBTQ people, and heterosexual people) as enacting prejudice and as targeted by prejudice when previously reporting their prejudice expectations from 0 (Not thinking of them at all) to 100 (Exclusively thinking about them).

# Results

In the subsequent analyses, we first test whether the effects of condition on expectations of prejudice and environmental behavior replicate previous studies (see Table 6 for descriptive and inferential statistics; see appendix for tests of moderation by group membership We then test whether these signals are perceived as a *response to a problem,* and whether this mechanism explains prejudice expectations. Next, we test whether prejudice expectations in outside spaces (i.e., on campus and in US, the same mediator from Studies 1 and 2) mediate support for movements, policies, and organizations related to diversity, equity, and inclusion. Finally, we test what groups participants report thinking of as the targets and enactors of prejudice when reporting prejudice expectations. Because there were no differences between the two safe space conditions for all relevant measures, they are collapsed into one condition in the subsequent analyses.

**Prejudice Expectations**

Replicating previous studies, participants in the safe space (vs. control) condition expected no more or less prejudice in the signaled space (the advisor's office) and expected more prejudice in the non-signaled spaces (on campus and in the United States).

**Expectations of Environmentally-harmful Behavior**

Replicating previous studies, participants in the control (vs. safe space) condition expected more sign-compliant behavior (i.e., less environmentally-harmful behavior) inside signaled space (the advisor's office), with this effect transferring to an adjacent space (i.e., campus) but not to a more distal space (i.e., the United States).

Table 6

*Expectations of environmentally-harmful behavior and prejudice from Study 4*

| | Control *M,* 95% CI | Safe Space *M*, 95% CI | *t* | *df* | *p* | *d* |
|---|---|---|---|---|---|---|
| Prejudice: Office | 17.97 [15.19, 20.75] | 18.47 [16.45, 20.48] | .29 | 819 | .78 | .02 |
| Prejudice: Campus | 30.19 [27.21, 33.16] | 38.76 [36.60, 40.93] | 4.58 | 820 | < .001 | .34 |
| Prejudice: US | 50.58 [47.49, 53.67] | 57.91 [55.67, 60.16] | 3.77 | 820 | < .001 | .28 |
| Eco-harmful: Office | 18.99 [16.19, 21.78] | 22.94 [20.91, 24.97] | 2.25 | 820 | .03 | .17 |
| Eco-harmful: Campus | 35.88 [33.00, 38.76] | 40.57 [38.58, 42.66] | 2.59 | 819 | .01 | .19 |
| Eco-harmful: US | 61.99 [59.11, 64.87] | 63.57 [61.48, 65.67] | .87 | 820 | .38 | .06 |

**Response to a Problem**

We tested whether the safe space sign was perceived as a response to a problem in two ways. First, we examined whether participants agreed that the sign was a response to a problem outside spaces by comparing their answers against a neutral scale midpoint. Second, we tested whether participants in the safe space (vs. eco-friendly) condition more strongly believed that the sign was a response to a problem in outside spaces.

Consistent with hypotheses, participants in the safe space condition agreed that the sign was a response to a problem on campus and in the United States via testing response means against the scale midpoint of "neither agree nor disagree" ($M_{SS-campus} = 4.87$, $t(534) = 13.62$, $p < .001$, $d = .59$; $M_{SS-US} = 4.95$, $t(534) = 14.08$, $p < .001$, $d = .61$). Participants in the eco-friendly condition also agreed that the Eco-Friendly sign was a response to a problem on campus and in the United States via testing response means against the scale midpoint ($M_{Eco-campus} = 4.57$, $t(282) = 6.63$, $p < .001$, $d = .40$; $M_{Eco-US} = 5.02$, $t(279) = 14.08$, $p < .001$, $d = .76$).

In addition, participants in safe space (vs. the eco-friendly) condition rated sign as more of a response to a problem on campus ($t(816) = 2.82$, $p = .005$, $d = .21$), but as no more of a response to a problem in the United States, $t(813) = .66$, $p = .51$, $d = .05$. This supports the idea

that identity safety signals are perceived as a response to a problem in outside spaces, and that compared to other signals, they are perceived as a more of a response to a problem in local outside spaces (but not as more of a response to a problem at the national level).

**Response to a Problem as Mechanism for Prejudice Expectations**

We predicted that believing the Safe Space sign was a response to a problem in outside spaces would predict prejudice expectations in those spaces. Supporting this, believing the safe space sign was a response to a problem on campus predicted increased prejudice expectations on campus, $r(533) = .36$, $p < .001$; likewise, believing the sign was a response to a problem in the United States predicted increased prejudice expectations in the United States, $r(533) = .32$, $p < .001$. For participants in the control condition, believing the Eco-friendly sign was a response to a problem on campus also predicted expectations of environmentally-harmful behavior on campus ($r(280) = .19$, $p = .002$) and believing the sign was a response to a problem in the United States predicted expectations of environmentally-harmful behavior in the United States, $r(278) = .17$, $p = .004$.

It is possible, however, that the relationship between seeing a sign as a response to a problem and expectations of a problem differs by sign type. For instance, the link between believing "Safe Space" signs are a response to prejudice and prejudice expectations might be stronger than the link between believing an "Eco-Friendly" sign is a response to environmentally-harmful behavior and expectations of environmentally-harmful behavior. If so, this would help explain why the safe space sign elicits problem expectations more so than the Eco-friendly sign. To examine this, we created a variable for expectations of the sign-relevant problem in outside spaces (i.e., composite of prejudice expectations on campus and United States for those in the Safe Space condition and composite of expectations of environmentally-harmful

behavior on campus and United States for those in the control condition). A moderated regression showed that the relationship between believing that the sign was a response to a problem in outside spaces and expectations of the problem in outside spaces was stronger in the Safe Space ($\beta = 7.82$, 95% CI [6.38, 9.25], SE = .73, $t(814) = 10.69$, $p < .001$) versus control condition ($\beta = 4.95$ 95% CI [2.70, 7.20], SE = 1.15, $t(814) = 4.32$, $p = <.001$), $\beta_{interaction} = 2.87$, 95% CI [.20, 5.54], SE = 1.36, $t(814) = 2.11$, $p = .04$. See Figure 4.

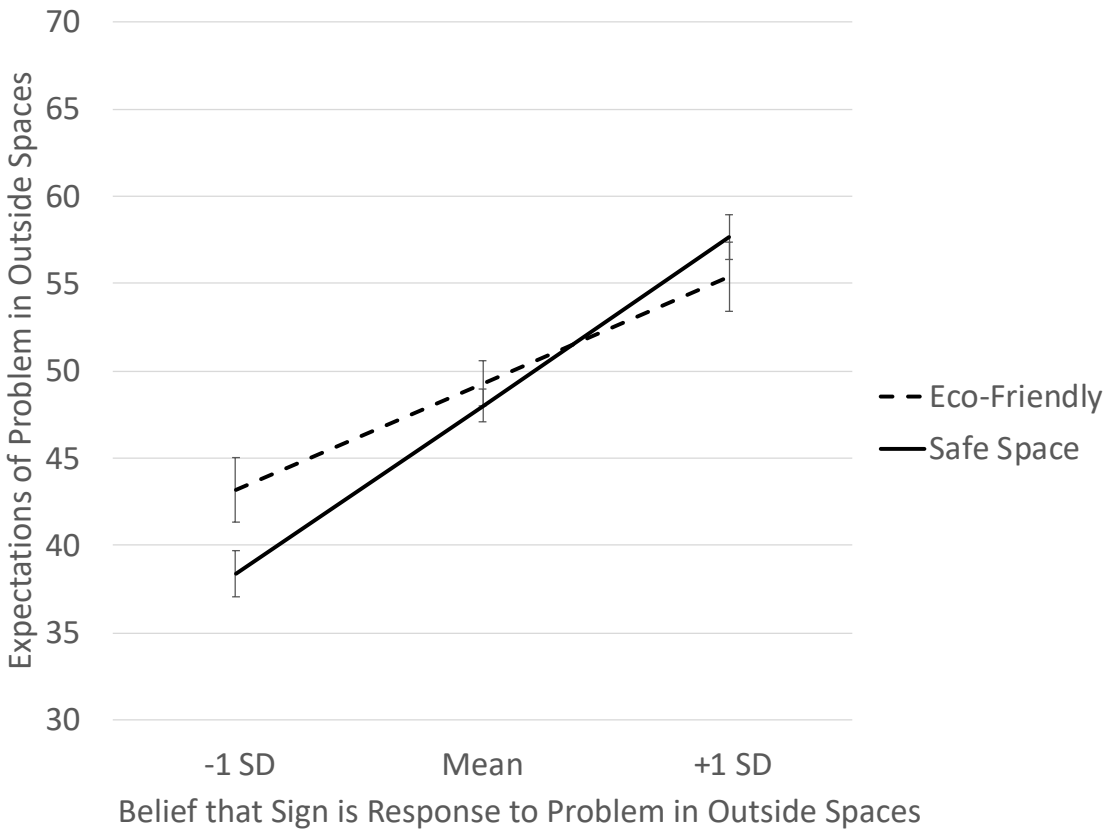



**Fig. 4.** Relationship between believing a sign is a response to a problem in outside spaces and expectations of the sign-relevant problem in outside spaces, moderated by condition (Study 4). Error bars represent +/- 1 Standard Error.

**Support for Movements, Policies, and Organizations Related to Diversity, Equity, and Inclusion**

Consistent with previous studies, safe space (vs. control) had positive indirect effects on support for well-known policies/movements and donations. The total effects of condition on support for well-known policies/movements and donations were in the right direction, but neither effect was statistically significant. See Figure 5 for mediation model and statistics.

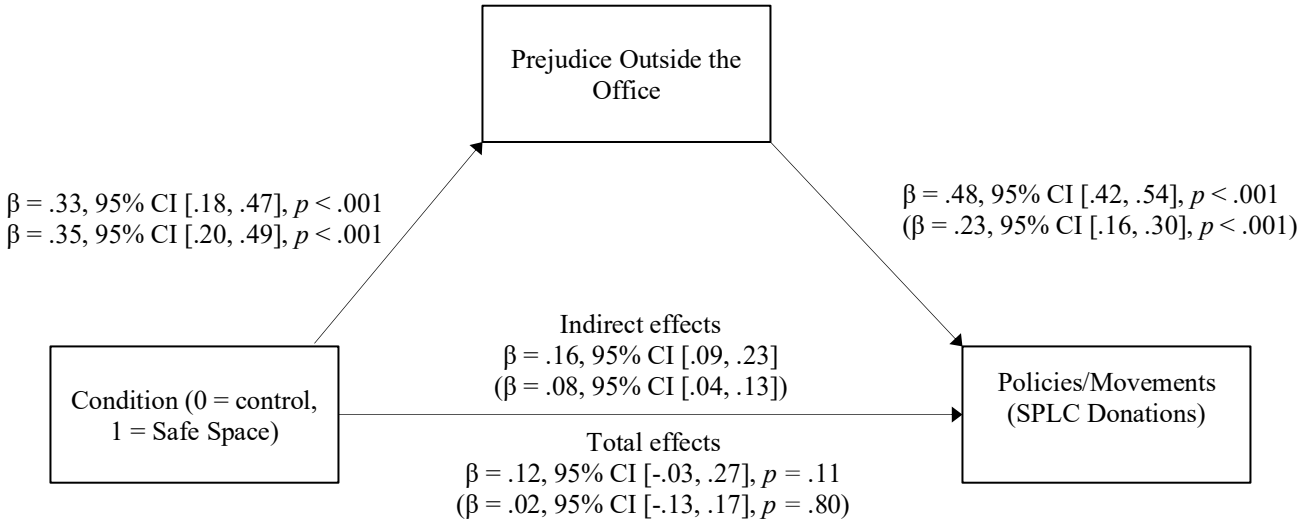
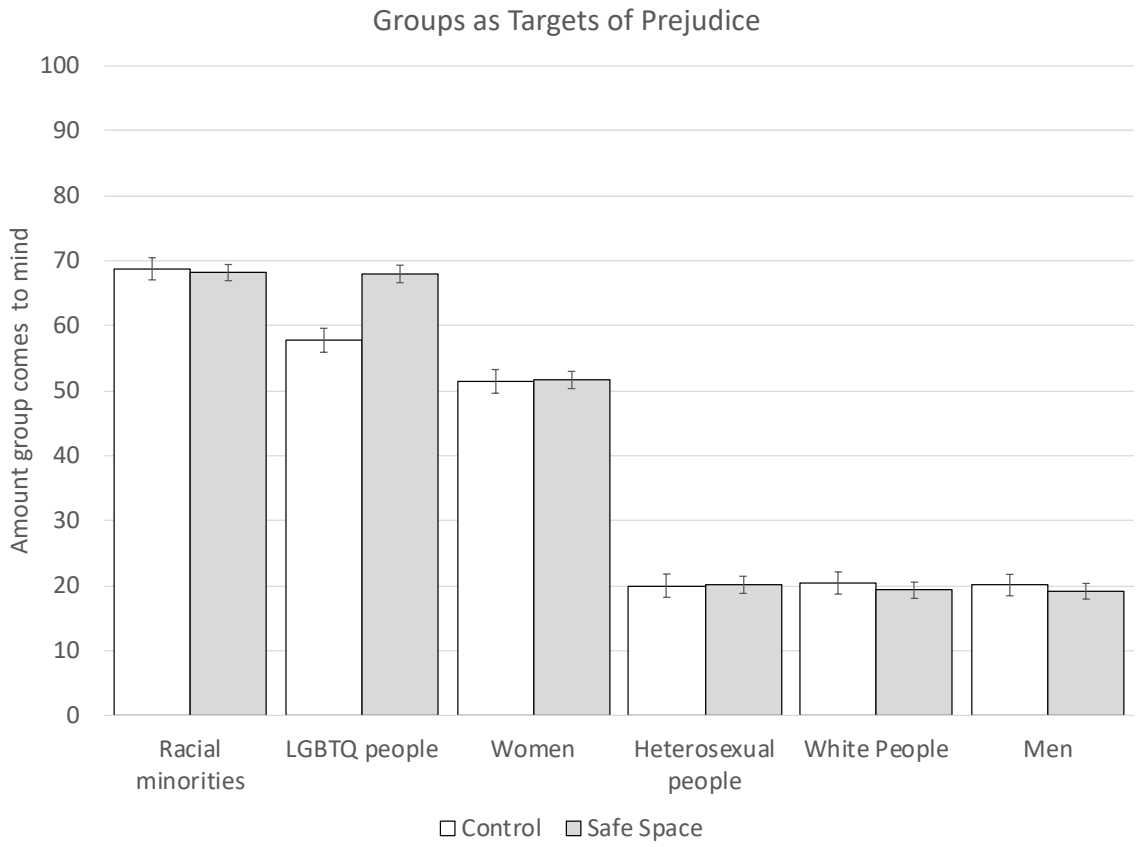



**Fig 5.** Effect of Safe Space cue on support for well-known DEI movements/policies and donations to the Southern Poverty Law Center, as mediated by prejudice expectations outside the office (Study 4).

### Targets and Enactors of Prejudice

Across studies, participants only reported the amount of prejudice they expected in various environments, but we never asked participants which groups they were thinking of as experiencing or enacting this prejudice. Consistent with predictions, participants reported imagining historically marginalized groups (i.e., racial minorities, women, LGBTQ people) as

the targets of prejudice and historically privileged groups (i.e., White people, men, heterosexual people) as the enactors of prejudice. See Figure 6 and 7.[2]




[2] We also analyzed whether the groups participants were thinking of as the targets and enactors of prejudice varied across condition. The only conditional difference was participants thinking of LGBTQ people more in the safe space (vs. control) condition, $F(1, 772) = 20.02$, $p < .001$, partial $\eta^2 = .025$.

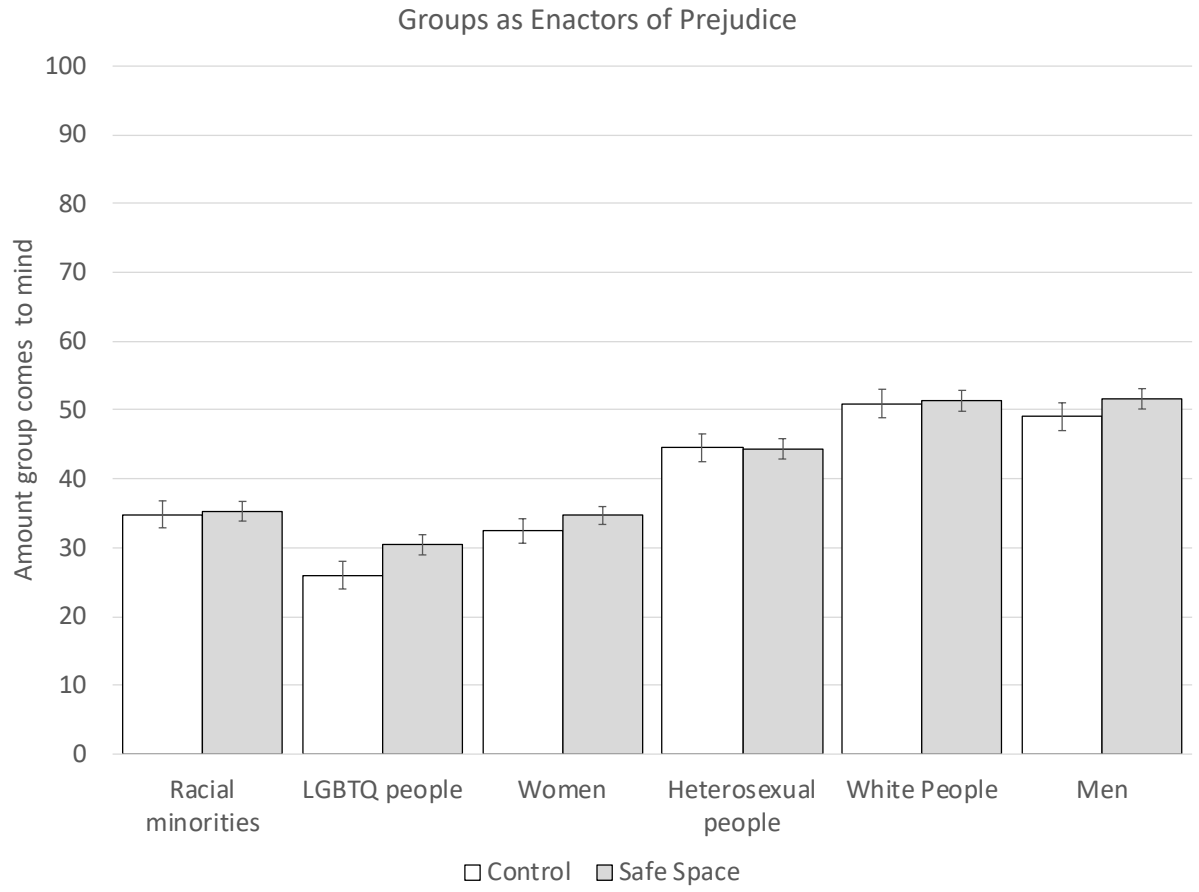



**Fig. 6 and 7.** Bar graphs depicting the degree to which different social identity groups came to mind as targets (Fig 6.) and enactors (Fig. 7) of prejudice when answering questions about prejudice expectations in Study 4. Error bars represent +/- 1 Standard Error.

## Discussion

Replicating previous studies, participants in the safe space (vs. control) condition expected more prejudice in outside spaces (campus and the United States), but no more or less prejudice in the signaled space. Also replicating previous studies, the safe space sign worked differently than a control sign. Unlike the safe space sign, the eco-friendly sign reduced expectations of the sign-relevant problem (i.e., environmentally-harmful behavior) in the signaled space and in an adjacent location (i.e., campus), but did not affect expectations of the problem in outside locations (e.g., the United States).

We reasoned that identity safety signals might increase prejudice expectations by signaling the presence of a problem (i.e., the *response to a problem* hypothesis). Consistent with this hypothesis, participants in the Safe Space condition agreed with the fact that the sign was a response to a problem on campus and in the United States, and believing that the sign was a response to a problem in those environments was a strong predictor of subsequent prejudice expectations in those environments. In addition, participants generally perceived the eco-friendly sign as a response to a problem to a lesser degree than they saw the Safe Space sign as a response to a problem. Moreover, the relationship between seeing the Safe Space sign as a response to a problem and subsequent prejudice expectations was stronger than the relationship between seeing the eco-friendly sign as a response to a problem and subsequent expectations of environmentally-harmful behavior. Together, this suggests that one reason identity safety signals increase prejudice expectations in other locations is that they signal an existing problem, to which the signal is a response. Furthermore, identity safety signals elicit this inference more than other environmental cues.

Contrary to our predictions, however, a subtle change in the definition of "safe spaces" did not reduce prejudice expectations or the extent to which people thought the sign was responding to a problem. It is possible that the term "safe space" has a strong connotation that cannot be shifted through subtle changes in definition; it is also possible that the intentional and explicit signals of identity safety, more broadly (regardless of framing), are generally closely linked with the problem behavior in people's minds, meaning that other identity signals meeting these criteria will elicit increased problem expectations.

Finally, Study 4 extended the previous three studies by demonstrating that when participants have been answering questions about prejudice expectations in the present research,

they by and large have been thinking about traditionally marginalized groups as the targets of this prejudice and historically privileged groups as the groups committing these acts of prejudice. Study 4 also echoed the findings of the prior studies by showing that exposure to identity safety signals have consistent indirect effects (but inconsistent total effects) on increased support toward for diversity, equity, and inclusion. In Study 4 this was demonstrated by way of indirect effects for elevated attitudes towards movements/policies supporting diversity, equity, and inclusion, as well as for increased donations to the Southern Poverty Law Center.

## GENERAL DISCUSSION

People are constantly making sense of their social environments, especially regarding potential threats. Through this lens, people are likely to use institutional identity safety signals to make inferences about prejudice in their broader social environments. Prior research, however, has not addressed how these signals affect how people think and feel about other, non-signaled environments. The present research addressed this gap by focusing on three goals.

The first goal was to test if institutional identity safety signals affected prejudice expectations inside and outside the signaled location. Across all studies, identity safety signals increased prejudice expectations in non-signaled environments, but more weakly for more distant environments. Identity safety signals, however, did not reduce (and sometimes marginally increased) prejudice expectations inside the signaled space. For instance, participants reading about a safe space (vs. a prejudice-irrelevant) signs in an advisor's office expected no more or less prejudice in the office, and participants reading about a company commitment to DEI (vs. to the environment) expected marginally more prejudice in the company. Finally, participants used identity safety signals differently from other institutional signals to inform expectations of their social environments. For instance, signals of eco-friendliness *reduced* people's expectations of

environmentally-harmful behavior in signaled space and in an adjacent space, but had no effect on expectations of environmentally-harmful in the United States.

In addition, we explored whether the effects of safety signals on prejudice expectations were moderated by group membership. When contrasting historically privileged and marginalized groups for gender, race, and sexual orientation, we found no evidence of moderation. This is somewhat surprising—because prejudice concerns are more chronically active for participants from historically marginalized (vs. privileged) groups (Mendoza-Denton et al., 2002), participants from historically marginalized (vs. privileged) groups could have been less influenced by safety signals that bring to prejudice to mind. Nonetheless, the influence of elevated prejudice expectations could vary across groups, such as increasing stress more for groups likely to experience prejudice. Future research should continue to explore how these signals are perceived by different groups, especially when the identity safety signals are group-specific. Recent research by Cundiff and colleagues (2018), for instance, suggests that diversity initiatives targeted at women (vs. to all people) lead both men and women to feel more concern about being treated unfairly in the signaling organization, despite the fact that both types of initiatives effectively signal a commitment to diversity.

The second aim was to determine the process through which signals to identity safety affected prejudice expectations in non-signaled spaces. One possibility was that identity safety signals would reduce prejudice expectations in the signaled space, and that these effects would transfer to other spaces (i.e., *the transferability hypothesis*). The hypothesis was ruled out because identity safety signals often did not affect prejudice expectations inside the signaled space and because they typically increased prejudice expectations in outside spaces. We tested two other hypotheses explaining increased prejudice expectations in outside spaces. The

*atypicality hypothesis* posited that environments with identity safety signals felt atypical, making non-signaled environments seem comparatively unsafe. This hypothesis was not supported, however—even when safety signals were typical of the broader environment, participants expected more prejudice in the broader environment (Study 2). The other possibility involved participants seeing identity safety signals as a response to a problem, thus inferring prejudice from the signal. Study 4 supported this: participants indicated that they saw the "Safe Space" sign as a response to a problem in outside spaces, which predicted increased prejudice expectations in outside spaces.

The third goal of the present research was to explore downstream consequences of increased prejudice expectations. Identity safety signals had indirect effects via increased prejudice expectations in non-signaled spaces on increased perceptions of prejudice (Study 1), intentions to confront prejudice and egalitarian attitudes (Studies 2 and 3), attitudes toward movements/policies supporting diversity, equity, and inclusion (Studies 3 and 4), and donations to an anti-prejudice organization (Study 4). At times, reading about identity safety signals directly increased these outcomes: in Study 2 our manipulation increased intentions to confront prejudice and egalitarian attitudes and in Study 4 it produced a marginal increase in support for movements/policies supporting diversity, equity, and inclusion. More often, however, reading about identity safety signals did directly influence these outcomes. This is consistent with other research showing that environmental nudges often have indirect effects on attitudes and behavior, absent direct effects (Derricks & Earl, 2019; Lewis & Earl, 2018). Moreover, it is possible that the observed effects of identity safety signals could amplify as they play out over time in a real-world context (Kenthirarajah & Walton, 2015). Together, these findings

demonstrate a number of ways that prejudice expectations can be beneficial for a society aiming to reduce prejudice and increase diversity, equity, and inclusion.

## Theoretical Implications

### Identity Safety Signals Transferring Across New Dimensions

The current studies are the first to show how organizational safety cues transfer across location. We focused on this dimension because people live their day-to-day in a multitude of locations—going inside and outside of spaces that are signaled to be safe and those that are not—motivating research into how these signals affect people's expectations of threat in the non-signaled spaces. Prior work has demonstrated that institutional signals of identity safety can transfer across groups (Chaney et al., 2016). Specifically, Chaney and colleagues (2016) showed that compared to a control condition, organizational signals of identity safety for racial minorities led women to perceive the organization's manager as lower in social dominance orientation, and in turn, led women to feel safer overall.

Here, the present research shows that "cue transfer" can operate for a different dimension (i.e.., across locations, rather than social identities) and through different processes (i.e., through inferences made about the impetus for these signals, rather than inferences about the beliefs of authority figures in the organizations). Taken together, the present research suggests that cues about identity safety may not transfer in the same ways for different dimensions. Future research should consider how cues to identity threat and safety transfer across other dimensions, such as time (e.g., expectations of identity safety and threat in the organization in the future) or other types of threat (e.g., expectations of identity-irrelevant threats based on signals of identity safety).

### Identity Safety Signals Eliciting Threat

Second, the present research demonstrates novel consequences following from perceiving institutional identity safety signals. Most research on these signals has shown their positive effects for identity safety in the signaling organization, such as trust and comfort within an organization (Purdie-Vaughns et al., 2008), feeling accepted by others in the organization (Meeussen, Otten, & Phalet, 2014), and increased perceptions of fairness (Chaney et al., 2016). On the other hand, other research has shown various ways in which these signals can elicit threat, such as affirmative action policies leading intended beneficiaries to question the merits of their success (Leibbrandt & List, 2018; Major et al., 1994) and to worry about being mistreated (Heilman & Alcott, 2001). The present research fits into this existing research by suggesting that certain types of signals (i.e., those that are intentional and explicit) can be perceived as a response to a problem, and in turn increase one kind of threat (i.e., prejudice expectations) outside the signaled location.

In addition, contrary to our predictions, the present research showed that identity safety signals do not necessarily reduce threat expectations *inside* the signaled space. This supports previous research demonstrating that identity safety signals do not always have the effects that people might logically expect them to have, and that instead, the effects of these signals should be empirically tested (Caleo & Heilman, 2019; Hideg & Wilson, 2020; Pietri et al., 2019; for a review, see Leslie, 2019). There are a number of reasons that might explain why identity safety signals may have failed to reduce prejudice expectations. When the advisor's office was the signaled space, for instance, there may have been a floor effect stemming from people generally expecting little prejudice in the space, regardless of condition. In other words, participants may not expect prototypical college advisors to act in a prejudiced manner, and may not expect the interactions that occur in their offices to feature prejudice, regardless of the presence of an

identity safety signal. Future research could systematically test whether this null effect is a function of the type of individual involved (e.g., a college advisor), of the setting (e.g., a formal educational environment), or number of individuals involved (e.g., people may not want to attribute prejudice to individual actors).

Speaking to this latter possibility, when the broader institution was a signaled space (i.e., the company in Study 3 and arguably the campus in Study 2 where safe space signs were prevalent), there was suggestive evidence that identity safety signals increased prejudice expectations inside the signaled space. This would be consistent with the idea that signals of identity safety at the institutional level may also be perceived as a response to a problem inside the institution itself. Future research should more systematically test how the size—both in terms of physical space and the number of people that inhabit the environment—of signaled and non-signaled spaces modulate the effects of identity safety signals on threat expectations in those signaled and non-signaled spaces.

Future research should continue to focus on the types of identity safety signals that elicit threat concerns. For instance, it is possible that intentional and explicit signals of identity safety are perceived as a response to a problem, whereas other signals of identity safety (e.g., representation of traditionally underrepresented groups in positions of power) are not perceived as institutional efforts to explicitly signal safety, and thus are not perceived as a response to a problem. In a similar vein, future research should explore whether the effects of identity safety signals depend on the details about how people are exposed to these signals. For instance, in the present research, participants read about hypothetical scenarios that involved protagonists coming across the signals in person. Follow-up studies could place participants in situations that are comparable to those protagonists, which would allow for better external validity and the

ability to test how people's prejudice expectations change over time (e.g., how long-lasting are people's changes in prejudice expectations?). One possibility, for instance, is that when there are more identity safety signals in the environment, people are exposed to these signals more over time, resulting in elevated prejudice expectations that solidify into stable beliefs over time.

**Implications for Normative Theory**

The various effects of institutional identity safety signals on prejudice expectations can be analyzed through a social norms framework, offering novel contributions to normative theory and new avenues for future research. In the present research, we provided institutional identity safety signals communicating that prejudice *should not* happen inside the signaled space (i.e., an *injunctive* social norm), and we subsequently measured people's expectations of whether prejudice actually occurs (i.e., a *descriptive* social norm) (Cialdini, Reno, & Kallgren, 1990; Cialdini, Reno, & Kallgren, 1991).

Prior research has typically shown a *congruent* relationship between injunctive and descriptive norms (Brauer & Chaurand, 2010; Cialdini et al., 2003; Eriksson, Strimling, & Coultas, 2015; Thøgersen, 2008). For instance, in a study examining public service announcements about the normativity of recycling, participants who saw an advertisement approving of recycling (i.e., an injunctive norm) subsequently perceived recycling to be more prevalent (i.e., a descriptive norm), and vice versa (Cialdini et al., 2003). The patterns in our control conditions accord with the congruent relationship between descriptive and injunctive norms demonstrated in prior work. For instance, an "Eco-Friendly" sign increased expectations of environmentally-friendly behavior (Study 2) and decreased expectations of environmentally-harmful behavior (Study 4) inside signaled spaces.

Interestingly, however, the conditions featuring identity safety signals produced

*incongruent* relationships between injunctive and descriptive norms (at least for the descriptive norms outside the signaled space). In these conditions, messages communicating an anti-prejudice injunctive norm *increased* the perception that prejudice was descriptively normative. Fortunately, the increased descriptive norm of prejudice was associated with increased motivation to *combat* prejudice. This result also stands in contrast to other work in normative theory, which often shows that increasing perceptions of problematic descriptive norms can increase norm-consistent problematic behavior (Cialdini et al., 1990). Our participants' motivation to combat the descriptive norm (i.e., prejudice) could be due to an elevated anti-prejudice injunctive norm overwhelming the effect of the descriptive norm.

Future research should explore why some normative signals produce congruent injunctive-descriptive norm relationships and others produce incongruent injunctive-descriptive norm relationships. For instance, people may associate some topics with descriptive norms, and others more so with injunctive norms. That could explain, for instance, why an "Eco-Friendly" sign led to a congruent injunctive-descriptive norm relationship, and the "Sage Space" sign produced no such relationship for descriptive norms around prejudice inside the signaled space. Similarly, signals pertaining to some topics may evoke proscriptive injunctive norms (i.e., disapproved behaviors), whereas others may evoke more prescriptive injunctive norms (i.e., approved behaviors); if prescriptive and proscriptive norms have different relationships with corresponding descriptive norms, then signals evoking different kinds of injunctive norms (i.e. prescriptive vs. proscriptive) could have divergent effects on perceptions of descriptive norms. Future research should 1) more precisely measure the specific normative associations with different kinds of signals, and 2) more precisely manipulate which types of norms are being communicated.

### Novel Consequences of Prejudice Expectations

Unlike most research on prejudice expectations, the present research highlights both costs (e.g., perceiving prejudice) and benefits (e.g., support for movements and policies that promote diversity, equity, and inclusion) of expecting prejudice. Although prior research has conceptually examined the relationship between awareness of prejudice and motivations to combat prejudice (Mallett et al., 2008; Paluck, 2011) the present research is the first, as far as we can tell, to directly demonstrate how increasing prejudice expectations can result in more favorable attitudes toward DEI, intentions to confront prejudice, positive attitudes toward movements and policies that address DEI, and donations to anti-prejudice organizations. More broadly, this suggests that other psychological research could benefit from considering other outcomes that follow for psychological states typically presumed to be negative. In addition, future research on identity safety signals should consider different kinds of outcomes—for instance, increased motivation to combat prejudice, ironically, may also lead to concerns about oneself appearing prejudiced, in turn interfering with smooth intergroup interaction (e.g., Goff, Steele, & Davies, 2008; Shelton, West, & Trail, 2010; Trawalter & Richeson, 2008).

### Safe Spaces as an Identity Safety Signal

Finally, the present research examined perceptions of a previously untested institutional identity safety signal: safe spaces. Safe spaces have been understudied in social psychology relative to the broader discussion surrounding them. By replicating our safe space findings with a company commitment to diversity, the present research also showed convergent validity as to how exposure to safe spaces and other intentional institutional identity safety signals affects prejudice expectations elsewhere. Future research on safe spaces, in particular, should consider testing their effects when they are targeted toward specific groups (e.g., LGBTQ people, the

group that developed the use of these spaces; Hanhardt, 2013; Kenney, 2001)—in these contexts, it is possible that people only expect more prejudice in outside spaces toward the targeted group, as that was the group for which a signal was created. In addition, future research on safe spaces should consider whether people expect other kinds of safety above and beyond identity safety—for instance, critics of safe spaces often wonder whether inhabitants of these spaces expect to be "safe" from their beliefs being questions or from encountering counter-attitudinal information (e.g., Ellison, 2016).

## Conclusion

Institutions are increasingly trying to create positive social climates. Although these efforts are often intended to signal identity safety and reduce expectations of social threats such as prejudice, these effects are often untested by scholars and institutions. Moreover, prior research has not considered how organizational identity safety signals influence expectations of other, non-signaled environments. The present research shows that in pursuit of signaling safety, institutions can also signal prejudice outside (and at times, even inside) those institutions. This demonstrates that creating inclusive climates is a complex issue, and that signaling safety is a strategy that carries a number of different consequences. As organizations continue to signal their egalitarian values and inclusive climate, they can carefully consider the positive and negative consequences we have outlined in the present research.

## References

Abrams, L. S., & Gibson, P. (2007). Teaching notes: Reframing multicultural education: Teaching white privilege in the social work curriculum. *Journal of Social Work Education*, *43*(1), 147-160

Arao, B., & Clemens, K. (2013). From safe spaces to brave spaces: A new dialogue around diversity and social justice. In L. M. Landreman (Ed.), The art of effective facilitation: Reflections from social justice educators (pp. 135–150). Sterling, VA: Stylus

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol 42.* (pp. 319–373). San Diego, CA: Academic Press

Bless, H., & Wänke, M. (2000). Can the same information be typical and atypical? How perceived typicality moderates assimilation and contrast in evaluative judgments. *Personality and Social Psychology Bulletin*, *26*, 306–314

Brandt, M. J., Chambers, J. R., Crawford, J. T., Wetherell, G., & Reyna, C. (2015). Bounded openness: The effect of openness to experience on intolerance is moderated by target group conventionality. *Journal of Personality and Social Psychology*, *109*(3), 549-568

Brief, A. P., Umphress, E. E., Dietz, J., Burrows, J. W., Butz, R. M., & Scholten, L. (2005). Community matters: Realistic group conflict theory and the impact of diversity. *Academy of Management Journal, 48*, 830–844

Brauer, M., & Chaurand, N. (2010). Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people's reactions to uncivil behaviors. *European Journal of Social Psychology*, *40*(3), 490-499

Caleo, S., & Heilman, M. E. (2019). What could go wrong? Some unintended consequences of gender bias interventions. *Archives of Scientific Psychology*, *7*(1), 71-80

Cialdini, R. B., Barrett, D. W., Bator, R., Demaine, L. J., Sagarin, B. J., Rhoads, K. V. L., & Winter, P. L. (2003). *Activating and aligning social norms for persuasive impact*. Unpublished manuscript, Arizona State University

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, *58*(6), 1015-1026

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology* (Vol. 24, pp. 201-234). Academic Press

Case, K. A. (2007). Raising white privilege awareness and reducing racial prejudice: Assessing diversity course effectiveness. *Teaching of Psychology*, *34*(4), 231-235

Chaney, K. E., & Sanchez, D. T. (2018). Gender-inclusive bathrooms signal fairness across identity dimensions. *Social Psychological and Personality Science*, *9*(2), 245-253

Chaney, K. E., Sanchez, D. T., & Remedios, J. D. (2016). Organizational Identity Safety Cue Transfers. *Personality and Social Psychology Bulletin*, *42*(11), 1564–1576

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*, 1045–1060

Cohen, L. L., & Swim, J. K. (1995). The differential impact of gender ratios on women and men: Tokenism, self-confidence, and expectations. *Personality and Social Psychology Bulletin*, *21*, 876–884

Cundiff, J. L., Ryuk, S., & Cech, K. (2018). Identity-safe or threatening? Perceptions of women-targeted diversity initiatives. *Group Processes & Intergroup Relations*, *21*(5), 745-766

Derricks, V., & Earl, A. (2019). Information targeting increases the weight of stigma: Leveraging relevance backfires when people feel judged. *Journal of Experimental Social Psychology*, *82*, 277-293

Donnelly, G. (2018, May 24). Starbucks Released Part of Its Diversity Training Curriculum. Retrieved January 7, 2019, from http://fortune.com/2018/05/24/starbucks-diversity-training

Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high- status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62,* 58-67

Ellison, J. (2016). Dear class of 2020 student. Retrieved from the University of Chicago website: https://news.uchicago.edu/sites/default/files/attachments/Dear_Class_of_2020_students.pdf

Emerson, K. T. U., & Murphy, M. C. (2014). Identity threat at work: How social identity threat and situational cues contribute to racial and ethnic disparities in the workplace. *Cultural Diversity and Ethnic Minority Psychology*, *20*(4), 508–520

Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, *129*, 59-69

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: stereotype threat and distance in interracial contexts. *Journal of personality and social psychology*, *94*(1), 91-107

Hanhardt, C. B. (2013). *Safe space: Gay neighborhood history and the politics of violence*. Duke University Press

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY

Heilman, M. E., & Alcott, V. B. (2001). What I think you think of me: Women's reactions to being viewed as beneficiaries of preferential selection. *Journal of Applied Psychology*, *86*(4), 574-582

Hideg, I., & Wilson, A. E. (2020). History backfires: Reminders of past injustices against women undermine support for workplace policies promoting women. *Organizational Behavior and Human Decision Processes*, *156*, 176-189

Hill, R. J. (2009). Incorporating queers: Blowback, backlash, and other forms of resistance to workplace diversity initiatives that support sexual minorities. *Advances in Developing Human Resources*, *11*(1), 37-53

Holley, L. C., & Steiner, S. (2005). Safe space: Student perspectives on classroom environment. *Journal of Social Work Education*, *41*(1), 49-64

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology*, 95, 796–805

Inzlicht, M., Kaiser, C. R., & Major, B. (2008). The face of chauvinism: How prejudice expectations shape perceptions of facial affect. *Journal of Experimental Social Psychology*, *44*(3), 758-766

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: ironic effects of organizational diversity structures. *Journal of personality and social psychology*, *104*(3), 504-519

Kaiser, C. R., Vick, S. B., & Major, B. (2006). Prejudice expectations moderate preconscious attention to cues that are threatening to social identity. *Psychological science*, *17*(4), 332-338

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. American Sociological Review, 71, 589–617. doi:10.1177/ 000312240607100404

Kenney, M. (2001). *Mapping gay LA: The intersection of place and politics*. Temple University Press

Kenthirarajah, D., & Walton, G. M. (2015). How brief social-psychological interventions can cause enduring effects. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 1-15

Kravitz, D. A., & Platania, J. (1993). Attitudes and beliefs about affirmative action: Effects of target and of respondent sex and ethnicity. *Journal of Applied Psychology*, 78, 928-938

Kumagai, A. K., & Lypson, M. L. (2009). Beyond cultural competence: critical consciousness, social justice, and multicultural education. *Academic Medicine*, *84*(6), 782-787

Leibbrandt, A., & List, J. A. (2018). *Do Equal Employment Opportunity Statements Backfire? Evidence From A Natural Field Experiment On Job-Entry Decisions* (No. w25035). National Bureau of Economic Research

Leslie, L. M. (2019). Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review*, *44*(3), 538-563

Lewis, N. A. Jr., & Earl, A. (2018). Seeing more and eating less: Effects of portion size granularity on the perception and regulation of food consumption. *Journal of personality and social psychology*, *114*(5), 786-803

Mallett, R. K., Huntsinger, J. R., Sinclair, S., & Swim, J. K. (2008). Seeing through their eyes: When majority group members take collective action on behalf of an outgroup. *Group Processes & Intergroup Relations*, *11*(4), 451-470

Major, B., Feinstein, J., & Crocker, J. (1994). Attributional ambiguity of affirmative action. *Basic and Applied Social Psychology*, *15*(1-2), 113-141

Major, B., Mendes, W. B., & Dovidio, J. F. (2013). Intergroup relations and health disparities: A social psychological perspective. *Health Psychology*, *32*(5), 514-524

Meeussen, L., Otten, S., & Phalet, K. (2014). Managing diversity: How leaders' multiculturalism and colorblindness affect work group functioning. *Group Processes & Intergroup Relations*, *17*(5), 629–644

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: implications for African American students' college experience. *Journal of Personality and SocialPpsychology*, *83*(4), 896-918

Murphy, M., & Destin, M. (2016). Promoting Inclusion and Identity Safety to Support College Success. Retrieved from https://s3-us-west2.amazonaws.com/production.tcf.org/app/uploads/2016/05/18140602/TCF_PromotingInclusionandIndentity.pdf

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*, 879–885.

Niemann, Y. F., & Dovidio, J. F. (1998). Relationship of solo status, academic rank, and perceived distinctiveness to job satisfaction of racial/ethnic minorities. *Journal of Applied Psychology*, *83*(1), 55-71

Operario, D., & Fiske, S. T. (2001). Ethnic identity moderates perceptions of prejudice: Judgments of personal versus group discrimination and subtle versus blatant bias. *Personality and Social Psychology Bulletin*, *27*(5), 550-561

Paluck, E. L. (2011). Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, *47*(2), 350-358

Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Bailey, A. H., Moss-Racusin, C. A., & Handelsman, J. (2019). Addressing unintended consequences of gender diversity interventions on women's sense of belonging in STEM. *Sex Roles*, *80*(9-10), 527-547

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–630. https://doi.org/10.1037/0022-3514.94.4.615

Safe Space. (2016). Retrieved from https://www.google.com/search?q=safe+space

Sawyer, P. J., Major, B., Casad, B. J., Townsend, S. S., & Mendes, W. B. (2012). Discrimination and the stress response: psychological and physiological consequences of anticipating prejudice in interethnic interactions. *American Journal of Public Health*, *102*(5), 1020-1026

Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high- and low-status groups. *Personality and Social Psychology Bulletin*, *28*, 694–707

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*, 68–74

Shaikh, A. (2018, January 25). University creates student taskforce to address bias incidents. Retrieved January 7, 2019, from https://www.michigandaily.com/section/campus-life/university-creates-taskforce-address-bias-incidents

Shelton, J. N., Richeson, J. A., & Salvatore, J. (2005). Expecting to be the target of prejudice: Implications for interethnic interactions. *Personality and Social Psychology Bulletin*, *31*(9), 1189-2012

Shelton, J. N., West, T. V., & Trail, T. E. (2010). Concerns about appearing prejudiced: Implications for anxiety during daily interracial interactions. *Group Processes & Intergroup Relations*, *13*(3), 329-344

Shulevitz, Judith (March 21, 2015). "In College and Hiding From Scary Ideas". Op-ed. *The New York Times*. Retrieved December 23, 2015

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 34.* (pp. 379–440). San Diego, CA: Academic Press

Stewart, T. L., Latu, I. M., Branscombe, N. R., Phillips, N. L., & Denney, H. T. (2012). White privilege awareness and efficacy to reduce racial inequality improve White Americans' attitudes toward African Americans. *Journal of Social Issues*, *68*(1), 11-27

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American psychologist*, *62*(4), 271-286

Thøgersen, J. (2008). Social norms and cooperation in real-life social dilemmas. *Journal of Economic Psychology*, *29*(4), 458-472

Trawalter, S., & Richeson, J. A. (2008). Let's talk about race, baby! When Whites' and Blacks' interracial contact experiences diverge. *Journal of Experimental Social Psychology*, *44*(4), 1214-1217

# Appendix

## Study 1

## Method

### Materials

**Vignette.** In Studies 1, 2, and 4, participants read a vignette that involved a Safe Space sign or a control sign. In Study 1, the control sign was a "No Cell Phone Use" sign, whereas in Study 2 and 4 it was an "Eco-Friendly" sign. In addition, Study 2 used an extra sentence to communicate that the advisor's sign was prevalent on campus (this sentence is bracketed below).

> It is 7:30 am on a Monday morning as John's alarm clock in his college dorm begins to ring incessantly. John is a college student at a local university and like many of his peers his days tend to start pretty early. Like any other day he will attend his set schedule of classes, but today there will be a small addition to his list of things to do. For the past few weeks John has been struggling with his grades in several classes, including an important physics class that he will need in order to graduate. John has tried several things on his own in an attempt to improve his grades, but as of yet none seem to be working. As such, John scheduled an appointment to speak with his academic advisor about what his options are.
>
> Arriving at the advisor's office several minutes early, John takes a seat and begins to play a game on his phone while waiting patiently for his appointment. After a few minutes John becomes bored of his phone and begins to look around the room. Towards the left side of the room he notices a clock on the wall, a thermostat, and a sign with the words "Safe Space" / "No Cell Phone Use" / "Eco-Friendly" written across; on the right side he observes a storage cabinet, a set of neatly stacked papers, and a small poster. [As he sees the sign, he realizes that he saw the same sign at the gym, the cafeteria, and in the offices of the other professors he's met with this semester].

**Scenarios.** In Study 1, participants read 16 scenarios and indicated whether each scenario was a microaggression, motivated by implicit prejudice, and motivated by explicit prejudice.

These scenarios are below. For each scenario, in parenthesis, we indicate whether it was ambiguous/microaggression, blatant, or benign.

> Waiting on a bench outside of a building you notice a young man walk up to the entrance. You witness him open the door for himself, turn around briefly to notice that an African-American girl is walking towards the building, and proceed to let the door close behind him. (Ambiguous/Microaggression)
>
> Sitting in a classroom on your university's campus, your professor is giving a lecture on how recent discoveries in quantum physics will affect the development of artificial intelligence. You witness him ask a fairly complicated question and even though Rosie put her hand up first, the professor calls on a young man named Robert instead. (Ambiguous/Microaggression)
>
> After finishing up the lesson plan for the day your professor announces who received the best grades on the last exam. When the top scores are announced, you overhear a person say that Asians are always at the top of this class. (Ambiguous/Microaggression)
>
> Walking up one side of the street you happen to notice that a young woman is walking directly in front of you while a large black man is walking towards you both along the same side. When he's about 20 feet away you notice that the young woman crosses the street. (Ambiguous/Microaggression)
>
> Getting your nails done at your local nail salon you notice an older woman asking a young stylist a lot of questions about what her life in China was like. After a few minutes of listening in on this conversation, you hear the older woman remark about how good the young stylist's English is. (Ambiguous/Microaggression)
>
> Palio's is a new upscale restaurant down the street from your apartment. One day, you decide to go out for lunch there with a friend. While eating, you notice that the manager, a young Asian-American man, is chatting with several of his customers. You see one of the customers ask him, "By the way, what country are you from?" (Ambiguous/Microaggression)
>
> The Board of Directors at Clarkston Regional Hospital are in the midst of a hiring debate. A few board members are pushing for further interviewing with several women doctors, arguing that it has become a community and industry concern to account for gender diversity when hiring. Dr. Harland, the oldest member and chairman of the hiring committee, comments that he thinks that the most qualified person should get the job. (Ambiguous/Microaggression)
>
> You, Ben, and your friend Adrian, a twenty-four year old black man, are shopping in a store in the local mall. At first you and Ben are checking out shirts, while Adrian is in the pants section. After five minutes, the three of you meet up again. Adrian whispers to you and Ben that he's going to leave the store because one of the store employees had been

following him since he walked inside. Not having noticed anything himself, Ben asks Adrian if he was really sure he was being followed. (Ambiguous/Microaggression)

Sarah and Johnny, a homosexual man, have been friends for several years. One weekend they are shopping at the mall together when Sarah notices that a group of friends she knows from college are also there. Walking over to the group Sarah says hello and introduces Johnny to everyone as her "gay best friend." (Ambiguous/Microaggression)

Jessica is a junior in her third year of college and she has made many friends on campus. One night while hanging out in the living room of her sorority house with a few other girls, Jessica tells them that she is actually bisexual. One of the girls looks up from her phone and says, "How can you be queer though, since you have a boyfriend?" (Ambiguous/Microaggression)

Ethan and Allie are giving a presentation to their work colleagues on a new corporate sustainability initiative they have been tasked with developing. When Allie is using a graph to demonstrate the benefits of the initiative Ethan notices blank stares from the audience, jumps in, and says, "What Allie was trying to say was..." (Ambiguous/Microaggression)

Samuel, Adam, Debbie, and Catherine are working together on a group project for their college biology course. In order to complete the assignment, each group must present one copy of a written lab report with the correct information. Knowing that his handwriting is probably the worst out of the group, Adam turns to Debbie and asks if she would like to physically write the report. (Ambiguous/Microaggression)

After watching a YouTube video about one of President Obama's recent speeches you decide to scroll through the comments section. You notice one comment thread where an African-American person stated "Obama is the best president ever!" and another white YouTuber replied "You're an idiot, you and your kind should all go back to Africa." (Blatant)

While grabbing some seats at a bar with your friends you casually notice a man and a woman chatting a few seats down from you. After a few minutes, you overhear the woman say that she definitely doesn't want him to buy her a drink. The man then gets up and yells "Whatever, I wasn't interested in a slut like you anyway." (Blatant)

Andrew is walking down the hallway at the hospital where he is visiting his sick mother. On his way to her room he walks past a friendly African-American nurse who smiles and says hello to him. Andrew smiles back and says, "How are you doing?" (Benign)

Shopping at a clothing store you find a shirt you really like and get in line to buy it. As you wait in line you hear the male cashier tell the woman in front of you that she is going to receive $3.81 in change back. (Benign)

**Measures**

Below we list measures from Study 1 that were not reported or fully explicated in the main text. Some of these measures also occurred in subsequent studies. We note in parenthesis which studies the measure was included in.

**Manipulation check.** Participants answered, "What (if any) signs were on the wall in John's advisor's office? (Choose all that apply)" using the following options: No smoking, No Cell Phone Use, Safe Space, Quiet Study. In Studies 2 and 4, No Cell Phones was replaced with Eco-Friendly. (Study 1abc, Study 2, Study 4).

**Safe Space definition.** Participants gave a free response answer to the question, "In your opinion, based on your previous knowledge and past experiences, what does the word 'safe space' mean?" After giving this definition, participants answered subsequent questions around safe spaces that were defined as "space where everyone can feel comfortable about expressing their identity without fear of discrimination or attack." (Study 1abc, Study 2).

**Safe Space frequency.** Participants answered, "How often do you see 'safe spaces' in your life?" from 1 (I have never seen a safe space) to 10 (Many times a day), which each scale point labeled a frequency in between those values. (Study 1abc, Study 2).

**Safe Space context.** Participants gave a free response answer to the question, "Where and in what contexts have you seen 'safe spaces' on the University of Michigan's campus?" (Study 1b)

**Peer attitudes.** Participants answered, "How do you think students at the University of Michigan feel about 'safe spaces' in general?" using a sliding scale between 0 (very negatively) to 100 (very positively). (Study 1b)

**Peer attitudes (free response).** Participants gave a free response answer to the question, "How do you think students at the University of Michigan feel about 'safe spaces' in general?" (Study 1b)

**Attitudes toward political correctness.** Participants answered "How positively or negatively do you feel about a culture that values 'political correctness'?" using a scale from 1 (Very negative) to 7 (Very positive). (Study 1abc, Study 2).

**Beliefs about safe spaces as coddling and protective.** Participants rated their agreement with two statements—"Safe spaces 'coddle' people, hurting them in the long run" and "Safe spaces protect people, helping them in the long run"—using a scale from 1 (Strongly disagree) to 7 (Strongly agree). (Study 1abc, Study 2, Study 4).

**Political orientation.** Participants answered, "What best describes your political orientation?" using a scale from 1 (Liberal) to 7 (Conservative). (Study 1abc, Study 2, Study 3).

**Experience feeling triggered.** Participants answered, "How much have you experienced the feeling of being emotionally 'triggered' by something, according to your definition of what it means for someone to be 'triggered'?" using a scale from 1 (Never) to 5 (All of the time). (Study 1abc, Study 2, Study 3).

**Experience as target of prejudice.** Participants answered, "In your experiences, on average, how often have you experienced prejudice/bias from others?" using a scale from 1 "Never/almost never" to 7 (On a near daily basis). (Study 1abc, Study 2, Study 3).

**Gender.** Participants answered, "Which of the following terms describe your gender?" by selecting one of the following options: male, female, trans-male, trans-female, prefer not to answer, and if none of those terms describe you, please specify. (Study 1abc, Study 2, Study 3, Study 4).

**Importance of gender to identity.** Participants answered, "How important is your gender to your overall identity?" using a scale from 1 (Not at all important) to 7 (Extremely important). (Study 1abc, Study 2, Study 3).

**Age.** Participants answered "What is your age? (please enter a numerical value, e.g., 21)" by entering a number. (Study 1abc, Study 2, Study 3, Study 4).

**Race/ethnicity.** Participants answered, "Which of the following terms describe your race/ethnicity? Select all terms that apply to you" by selecting any number of the following: White or Caucasian, African American or Black, Hispanic or Latino/a, Native Hawaiian or Pacific Islander, Native American or American Indian, Asian or Asian American, South Asian or South Asian American, Middle Eastern American or Arab, Prefer not to respond, None of these options describe you (please specify below). (Study 1abc, Study 2, Study 3, Study 4).

**Importance of race/ethnicity to identity.** Participants answered, "How important is your race/ethnicity to your overall identity?" using a scale from 1 (Not at all important) to 7 (Extremely important). (Study 1abc, Study 2, Study 3).

**Sexual orientation.** Participants answered, "Which of the following terms describe your sexual orientation? Select all terms that apply to you" by selecting any number of the following: Heterosexual, Bisexual, Homosexual, Asexual, Queer, Prefer not to answer. (Study 1abc, Study 2, Study 3, Study 4).

**Importance of sexual orientation to identity.** Participants answered, "How important is your sexual orientation to your overall identity?" using a scale from 1 (Not at all important) to 7 (Extremely important). (Study 1abc, Study 2, Study 3).

**Education.** Participants answered "What is the highest education level you have achieved?" by selecting one of the following: Less then high school, High school, GED,

Bachelor's Degree, Associate's Degree, Master's Degree, Professional Degree (e.g., JD, MBA, MD, PharmD, PsyD, PhD), Other (please specify). (Study 1a and 1c, Study 2, Study 3, Study 4).

**Year in school.** Participants answered, "What year are you in school?" using one of five options: Freshman, Sophomore, Junior, Senior, Graduate Student. (Study 1b).

**Attention.** Participants answered, "In your honest opinion, how closely did you pay attention to the scenarios and questions in today's survey? Your answer to this question won't affect your compensation whatsoever, we just need to know to ensure the validity of our final data" from 1 (Not at all closely) to 5 (Extremely closely). (Study 1abc, Study 2, Study 3, Study 4).

**Suspicion.** Participants gave a free response answer to the question, "What do you think was being studied in this survey?" (Study 1abc, Study 2, Study 3, Study 4).

**Voting choice.** Participants answered, "In the most recent presidential election, who did you vote for?" by selecting Hilary Clinton, Donald Trump, Other Candidate, or Didn't Vote. (Study 1b).

**Candidate favorability.** Participants answered, "How favorable is your opinion of Donald Trump / Hillary Clinton?" from 0 (extremely unfavorable) to 100 (extremely favorable) for each candidate. (Study 1b).

**Prejudice concern post-Trump.** Participants answered, "Compared to before Donald Trump won the presidential election, how worried are you about prejudice being a problem in the United States?" from 1 (Much less worried) to 7 (Much more worried). (Study 1b).

**Safety post-Trump.** Participants answered, "Compared to before Donald Trump won the presidential election, do you feel more or less safe in general?" from 1 (Much less safe) to 7 (Much more safe). (Study 1b).

**Rights post-Trump.** Participants answered, “Compared to before Donald Trump won the presidential election, do you feel that your rights are more or less protected in general?” from 1 (Much less protected) to 7 (Much more protected). (Study 1b).

**Emotions post-Trump.** Participants answered, “Compared to before Donald Trump won the presidential election, how much do you generally feel each of the following emotions?” from 1 (Much less) to 7 (Much more) for: Anger, Joy, Fear, Disgust, Surprise, Sadness, Anxiety, Guilt, Excitement, Interest, Contempt. (Study 1b).

**Election thoughts.** Participants gave a free response answer to the prompt: Please write any other thoughts or feelings you have around this year's presidential election in the US that you would like to share. (Study 1b).

## Results

Supplementary Table 2 shows tests of how race, gender, and sexual orientation moderate the effect of condition in Study 1. Supplementary Table 3 shows tests of how continuous moderators (e.g., political orientation) moderate the effect of condition in Study 1. We include figures that represent any moderation by the continuous variables.

Table 7

Study 1 Moderation by Race, Gender, and Sexual Orientation

| | Non-White | | White | | Interaction statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 18.98 [15.55, 22.41] | 21.32 [19.95, 24.70] | 15.41 [13.22, 17.59] | 16.15 [13.92, 18.38] | 1033 | 0.3 | 0.58 | 0.000 |
| Prejudice on campus | 33.78 [30.02, 37.53] | 44.63 [40.93, 48.33] | 32.27 [29.87, 34.66] | 41.17 [38.73, 43.62] | 1033 | 0.4 | 0.54 | 0.000 |
| Prejudice in US | 60.23 [56.48, 63.97] | 64.87 [61.19, 68.55] | 52.64 [50.26, 55.02] | 58.64 [56.21, 61.07] | 1033 | 0.2 | 0.67 | 0.000 |
| Microaggression | 3.94 [3.75, 4.13] | 3.93 [3.74, 4.11] | 3.84 [3.72, 3.96] | 3.87 [3.75, 3.99] | 1033 | 0.1 | 0.77 | 0.000 |
| Implicit Bias | 4.33 [4.17, 4.50] | 4.30 [4.14, 4.47] | 4.26 [4.15, 4.36] | 4.23 [4.13, 4.34] | 1033 | 0.0 | 0.99 | 0.000 |
| Explicit Bias | 3.85 [3.69, 4.02] | 3.61 [3.45, 3.77] | 3.48 [3.37, 3.58] | 3.40 [3.29, 3.51] | 1033 | 1.4 | 0.23 | 0.001 |
| | Women | | Men | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 16.77 [14.22, 19.32] | 19.76 [17.16, 22.37] | 15.95 [13.26, 18.64] | 15.36 [12.67, 18.05] | 1024 | 1.8 | 0.18 | 0.002 |
| Prejudice on campus | 33.62 [30.87, 36.37] | 47.59 [44.79, 50.39] | 31.26 [28.37, 34.16] | 36.34 [33.44, 39.24] | 1024 | 9.5 | 0.002 | 0.009 |
| Prejudice in US | 58.10 [55.37, 60.83] | 66.41 [63.63, 69.20] | 51.27 [48.39, 54.15] | 53.82 [50.94, 56.70] | 1024 | 4.0 | 0.05 | 0.004 |
| Microaggression | 4.07 [3.93, 4.20] | 4.17 [4.03, 4.31] | 3.65 [3.51, 3.79] | 3.56 [3.42, 3.71] | 1024 | 1.7 | 0.19 | 0.002 |
| Implicit Bias | 4.47 [4.35, 4.59] | 4.51 [4.39 4.63] | 4.07 [3.94, 4.19] | 3.97 [3.84, 4.09] | 1024 | 1.3 | 0.26 | 0.001 |
| Explicit Bias | 3.70 [3.58, 3.82] | 3.61 [3.49, 3.74] | 3.48 [3.35, 3.61] | 3.31 [3.19, 3.44] | 1024 | 0.4 | 0.53 | 0.000 |
| | Non-heterosexual | | Heterosexual | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 15.99 [10.84, 21.13] | 15.46 [10.35, 20.56] | 16.37 [14.40, 18.35] | 18.00 [16.01, 19.99] | 1032 | 0.3 | 0.59 | 0.000 |
| Prejudice on campus | 33.57 [27.92, 39.21] | 43.63 [38.03, 49.24] | 32.54 [30.37, 34.71] | 42.02 [39.83, 44.21] | 1032 | 0.0 | 0.89 | 0.000 |
| Prejudice in US | 55.66 [50.03, 61.28] | 69.79 [64.21, 75.38] | 54.49 [52.33, 56.66] | 59.28 [57.10, 61.46] | 1032 | 4.7 | 0.03 | 0.004 |
| Microaggression | 3.71 [3.43, 3.99] | 4.32 [4.05, 4.60] | 3.89 [3.78, 4.00] | 3.81 [3.70, 3.92] | 1032 | 10.2 | 0.001 | 0.010 |
| Implicit Bias | 4.44 [4.19, 4.68] | 4.53 [4.28, 4.77] | 4.25 [4.16, 4.34] | 4.21 [4.12, 4.31] | 1032 | 0.5 | 0.49 | 0.000 |
| Explicit Bias | 3.37 [3.12, 3.62] | 3.53 [3.29, 3.78] | 3.62 [3.52, 3.71] | 3.46 [3.36, 3.55] | 1032 | 2.9 | 0.09 | 0.003 |

Table 8

Study 1 Moderation by Continuous Moderators

| | Interaction statistics - Protective Beliefs | | | |
|---|---|---|---|---|
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 1043 | 1.4 | 0.23 | 0.001 |
| Prejudice on campus | 1043 | 4.3 | 0.04 | 0.004 |
| Prejudice in US | 1043 | 0.6 | 0.43 | 0.001 |
| Microaggression | 1043 | 0.3 | 0.61 | 0.000 |
| Implicit Bias | 1043 | 0.6 | 0.44 | 0.001 |
| Explicit Bias | 1043 | 0.0 | 0.89 | 0.000 |
| | Interaction statistics - PC Culture Attitudes | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 1043 | 1.0 | 0.32 | 0.001 |
| Prejudice on campus | 1043 | 0.1 | 0.70 | 0.000 |
| Prejudice in US | 1043 | 0.4 | 0.54 | 0.000 |
| Microaggression | 1043 | 0.7 | 0.42 | 0.001 |
| Implicit Bias | 1043 | 0.4 | 0.51 | 0.000 |
| Explicit Bias | 1043 | 1.2 | 0.27 | 0.001 |
| | Interaction statistics - Political Orientation | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 1039 | 0.7 | 0.39 | 0.001 |
| Prejudice on campus | 1039 | 0.1 | 0.78 | 0.000 |
| Prejudice in US | 1039 | 2.3 | 0.13 | 0.002 |
| Microaggression | 1039 | 0.0 | 0.86 | 0.000 |
| Implicit Bias | 1039 | 0.8 | 0.36 | 0.001 |
| Explicit Bias | 1039 | 0.7 | 0.40 | 0.001 |
| | Interaction statistics - Prejudice Experience | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 1038 | 0.0 | 0.92 | 0.000 |
| Prejudice on campus | 1038 | 0.3 | 0.57 | 0.000 |
| Prejudice in US | 1038 | 1.6 | 0.21 | 0.002 |
| Microaggression | 1038 | 0.8 | 0.38 | 0.001 |
| Implicit Bias | 1038 | 0.2 | 0.63 | 0.000 |
| Explicit Bias | 1038 | 6.8 | 0.01 | 0.007 |

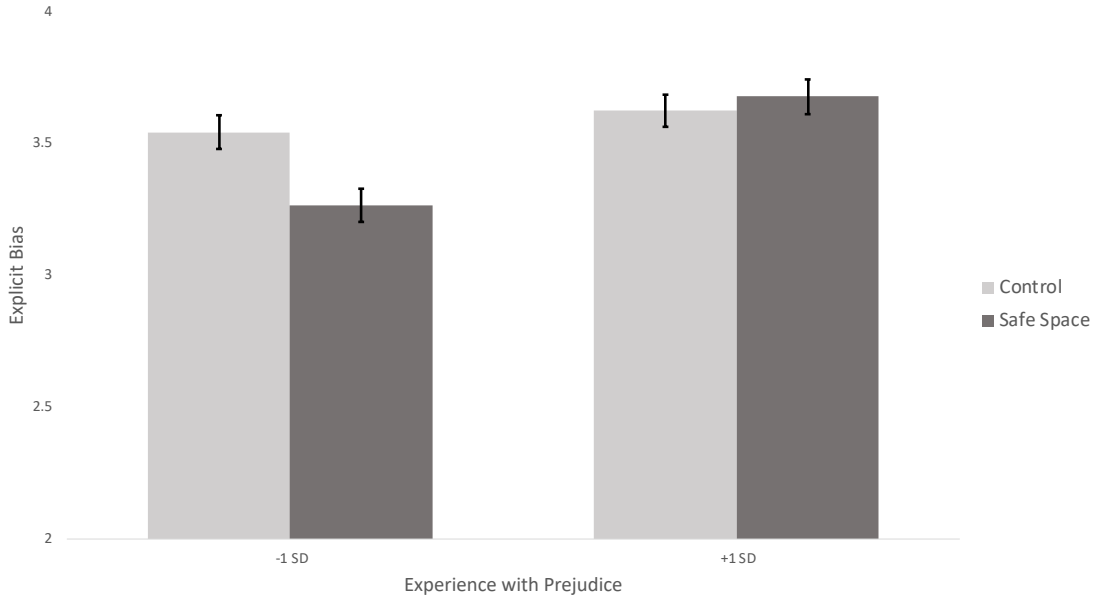



**Fig 8.** Study 1 interaction between condition and experience with prejudice for ratings of explicit bias in the scenarios with ambiguous prejudice. Error bars represent +/- 1 Standard Error.

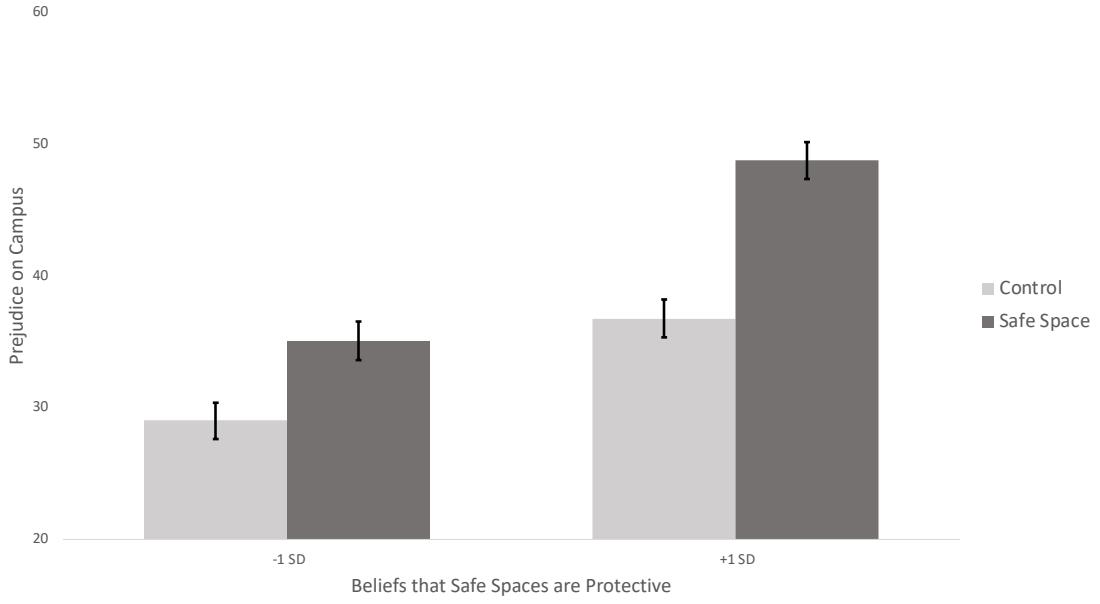



**Fig 9.** Study 1 interaction between condition and beliefs about safe spaces as protective (vs. coddling) for prejudice expectations on campus. Error bars represent +/- 1 Standard Error.

## Study 2

## Method

### Measures

Below we list Study 2 measures that were not reported or fully explicated in the main text, some of which also occurred in subsequent studies. We note in parenthesis which studies the measure was included in.

**Values safety compared to other schools.** Participants answered, “To what extent do you think John's college values a safe environment for students of all identities and backgrounds, as compared to other colleges?” from 1 (A lot less) to 7 (A lot more). (Study 2).

**Values environment compared to other schools.** Participants answered, “To what extent do you think John's college values environmentally-friendly behavior, as compared to other colleges?” from 1 (A lot less) to 7 (A lot more). (Study 2).

**Intentions to confront prejudice.** Participants rated their agreement from 1 (Strongly disagree) to 7 (Strongly agree) with each of the following statements: “If I witnessed an act of discrimination, I would see if the victim was okay,” “If I witnessed an act of discrimination, I would report it, if possible,” “If I witnessed an act of discrimination, I would intervene, if possible.” (Study 2, Study 3).

**Egalitarian attitudes.** Participants rated their agreement from 1 (Strongly disagree) to 7 (Strongly agree) with each of the following statements: “I would like to be exposed more to the perspectives of marginalized groups,” “Examining my own biases is important to me,” “I care about supporting movements that further the interests of marginalized groups,” “Being inclusive of people from different groups is important to me,” “I care about seeking out relationships with

people from different backgrounds," "Taking actions to increase diversity in institutions such as schools and companies is important." (Study 2, Study 3).

**Internal and external motivation to respond without prejudice.** Participants completed two five-item scales around their internal and external motivations to respond without prejudice (Plant & Devine, 1998). (Study 2, Study 3).

**Bias awareness.** Participants completed a four-item scale measuring the awareness of their own bias (Perry, Murphy, & Dovidio, 2015), which was adapted to be about awareness of bias toward all races, rather than just Blacks. (Study 2, Study 3).

## Results

Supplementary Table 4 shows tests of how race, gender, and sexual orientation moderate the effect of condition in Study 2. Supplementary Table 5 shows tests of how continuous moderators (e.g., political orientation) moderate the effect of condition in Study 2. We include figures that represent any moderation by the continuous variables.

Table 9

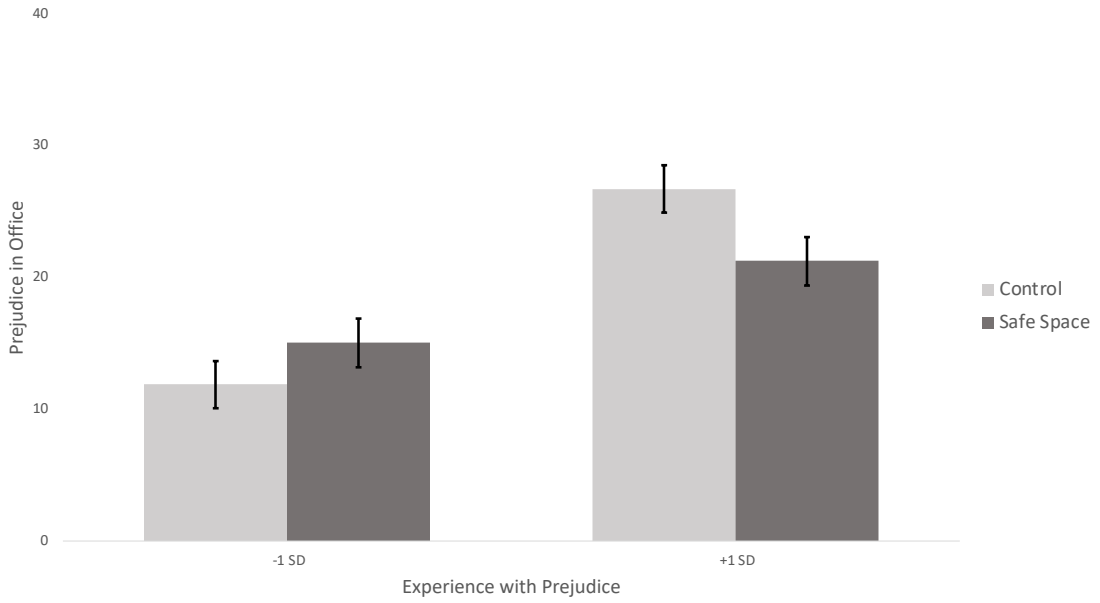
Study 2 Moderation by Race, Gender, and Sexual Orientation

| | Non-White | | White | | Interaction statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 20.53 [15.85, 25.20] | 24.19 [19.77, 28.60] | 18.71 [15.63, 21.79] | 14.93 [11.69, 18.16] | 589 | 3.5 | 0.06 | 0.006 |
| Prejudice on campus | 32.20 [27.24, 37.16] | 44.94 [40.26, 49.63] | 31.82 [28.55, 35.08] | 37.48 [34.05, 40.92] | 589 | 2.8 | 0.10 | 0.005 |
| Prejudice in US | 54.39 [49.30, 59.47] | 63.16 [58.36, 67.96] | 51.74 [48.40, 55.09] | 59.85 [56.34, 63.70] | 589 | 0.0 | 0.88 | 0.000 |
| Confront Prejudice | 5.48 [5.24, 5.72] | 5.69 [5.46, 5.91] | 4.45 [5.29. 5.61] | 5.62 [5.45, 5.79] | 588 | 0.0 | 0.88 | 0.000 |
| DEI attitudes | 5.27 [5.03, 5.52] | 5.54 [5.31, 5.77] | 5.11 [4.95, 5.28] | 5.30 [5.14, 5.47] | 588 | 0.1 | 0.71 | 0.000 |
| | Women | | Men | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 20.26 [16.67, 23.85] | 17.40 [14.07, 20.73] | 18.51 [14.70, 22.32] | 19.48 [15.14, 23.82] | 585 | 1.0 | 0.32 | 0.002 |
| Prejudice on campus | 33.32 [29.54, 37.10] | 41.27 [37.76, 44.79] | 30.80 [26.78, 34.82] | 37.94 [33.37, 42.52] | 585 | 0.0 | 0.84 | 0.000 |
| Prejudice in US | 57.18 [53.40, 60.96] | 65.41 [61.90, 68.92] | 47.36 [43.34, 51.37] | 53.51 [48.94, 58.08] | 585 | 0.3 | 0.61 | 0.000 |
| Confront Prejudice | 5.58 [5.40, 5.76] | 5.84 [5.68, 6.01] | 5.31 [5.12, 5.40] | 5.30 [5.08, 5.52] | 584 | 2.0 | 0.16 | 0.003 |
| DEI attitudes | 5.29 [5.10, 5.47] | 5.59 [5.42, 5.76] | 5.00 [4.81, 5.20] | 5.04 [4.81, 5.26] | 584 | 1.8 | 0.18 | 0.003 |
| | Non-heterosexual | | Heterosexual | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 24.14 [16.59, 31.69] | 19.48 [11.46, 27.51] | 18.37 [15.62, 21.11] | 18.00 [15.24, 20.77] | 588 | 0.5 | 0.47 | 0.001 |
| Prejudice on campus | 36.69 [28.70, 44.68] | 45.87 [37.38, 54.36] | 31.11 [28.21, 34.01] | 39.40 [36.48, 42.33] | 588 | 0.0 | 0.89 | 0.000 |
| Prejudice in US | 57.11 [48.94, 65.29] | 68.77 [60.09, 77.46] | 51.84 [48.87, 54.81] | 60.08 [57.09, 63.08] | 588 | 0.3 | 0.60 | 0.000 |
| Confront Prejudice | 5.57 [5.18, 5.96] | 5.98 [5.67, 6.39] | 5.44 [5.30, 5.58] | 5.60 [5.46, 5.75] | 587 | 0.7 | 0.42 | 0.001 |
| DEI attitudes | 5.41 [5.01, 5.80] | 6.04 [5.62, 6.45] | 5.12 [4.98, 5.27] | 5.31 [5.17, 5.45] | 587 | 2.0 | 0.15 | 0.003 |

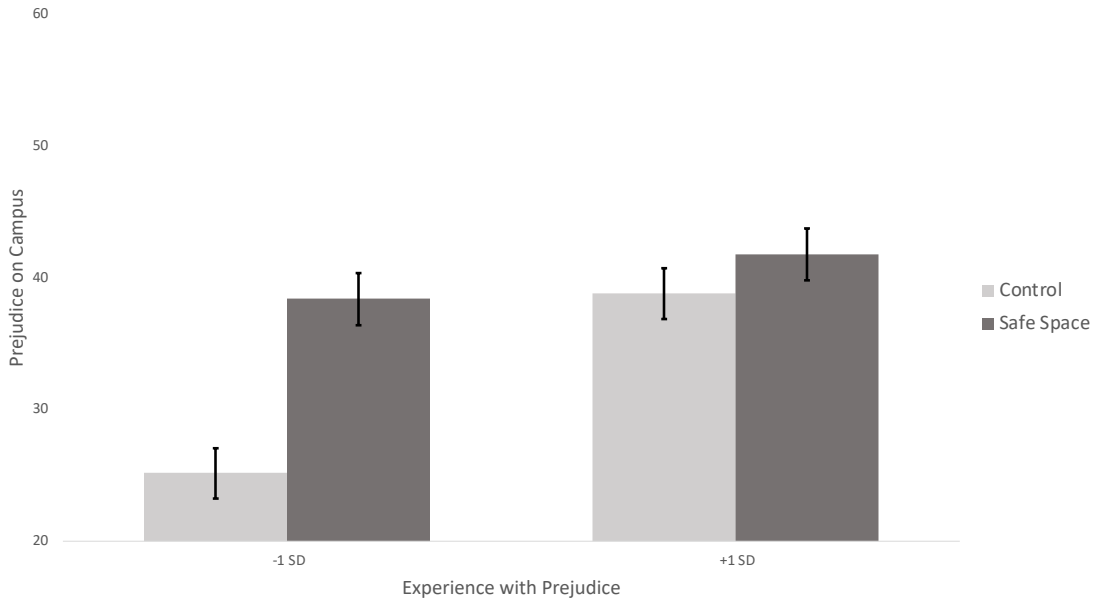
Table 10

Study 2 Moderation by Continuous Moderators

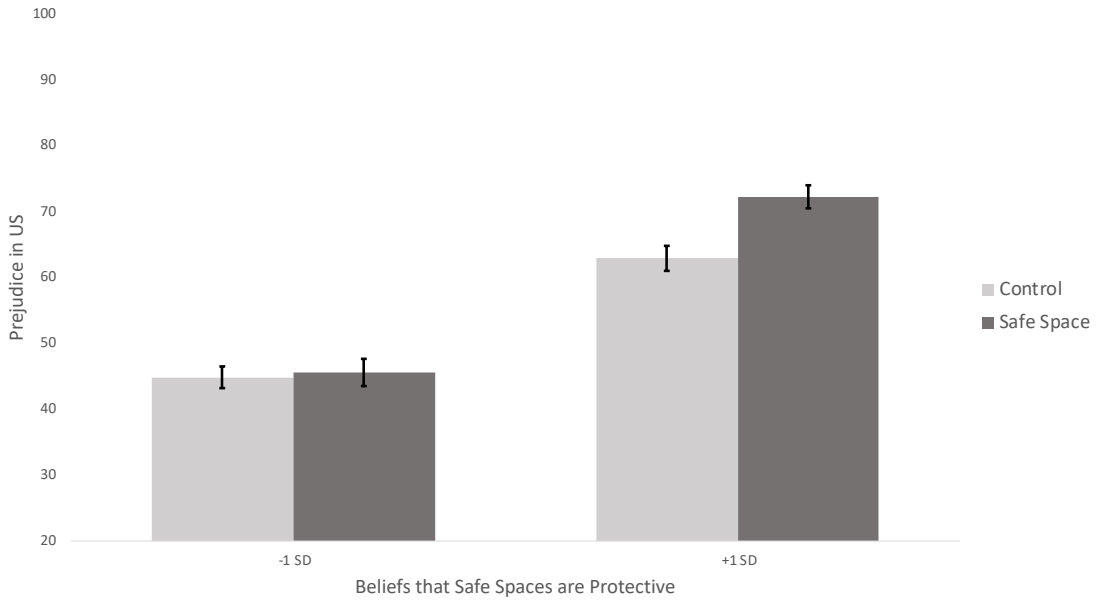
| | Interaction statistics - Protective Beliefs | | | |
|---|---|---|---|---|
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 589 | 0.1 | 0.77 | 0.000 |
| Prejudice on campus | 589 | 3.0 | 0.09 | 0.005 |
| Prejudice in US | 589 | 5.4 | 0.02 | 0.009 |
| Confront Prejudice | 588 | 1.5 | 0.23 | 0.003 |
| DEI attitudes | 588 | 2.0 | 0.15 | 0.003 |
| | Interaction statistics - PC Culture Attitudes | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 589 | 1.4 | 0.24 | 0.002 |
| Prejudice on campus | 589 | 1.1 | 0.29 | 0.002 |
| Prejudice in US | 589 | 2.2 | 0.14 | 0.004 |
| Confront Prejudice | 588 | 0.6 | 0.45 | 0.001 |
| DEI attitudes | 588 | 0.1 | 0.74 | 0.000 |
| | Interaction statistics - Political Orientation | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 588 | 0.6 | 0.45 | 0.001 |
| Prejudice on campus | 588 | 2.6 | 0.11 | 0.004 |
| Prejudice in US | 588 | 0.3 | 0.60 | 0.000 |
| Confront Prejudice | 587 | 3.7 | 0.05 | 0.006 |
| DEI attitudes | 587 | 5.2 | 0.02 | 0.009 |
| | Interaction statistics - Prejudice Experience | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 589 | 5.6 | 0.02 | 0.009 |
| Prejudice on campus | 589 | 6.9 | 0.01 | 0.012 |
| Prejudice in US | 589 | 2.6 | 0.11 | 0.004 |
| Confront Prejudice | 588 | 0.1 | 0.74 | 0.000 |
| DEI attitudes | 588 | 0.8 | 0.38 | 0.001 |



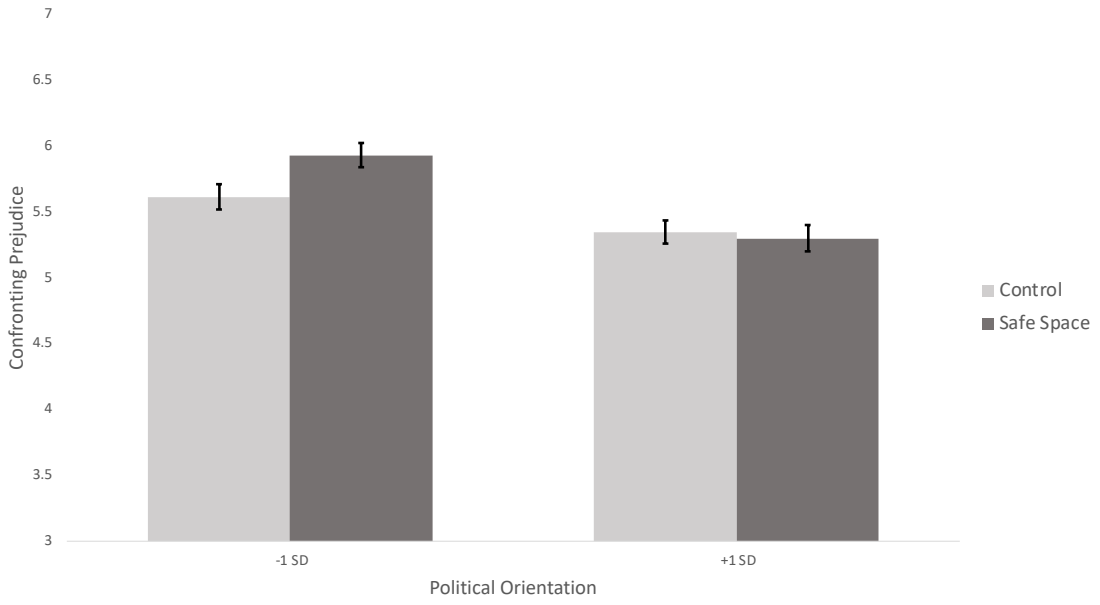

**Supplementary Fig 10.** Study 2 interaction between condition and experience with prejudice for prejudice expectations in the office. Error bars represent +/- 1 Standard Error.



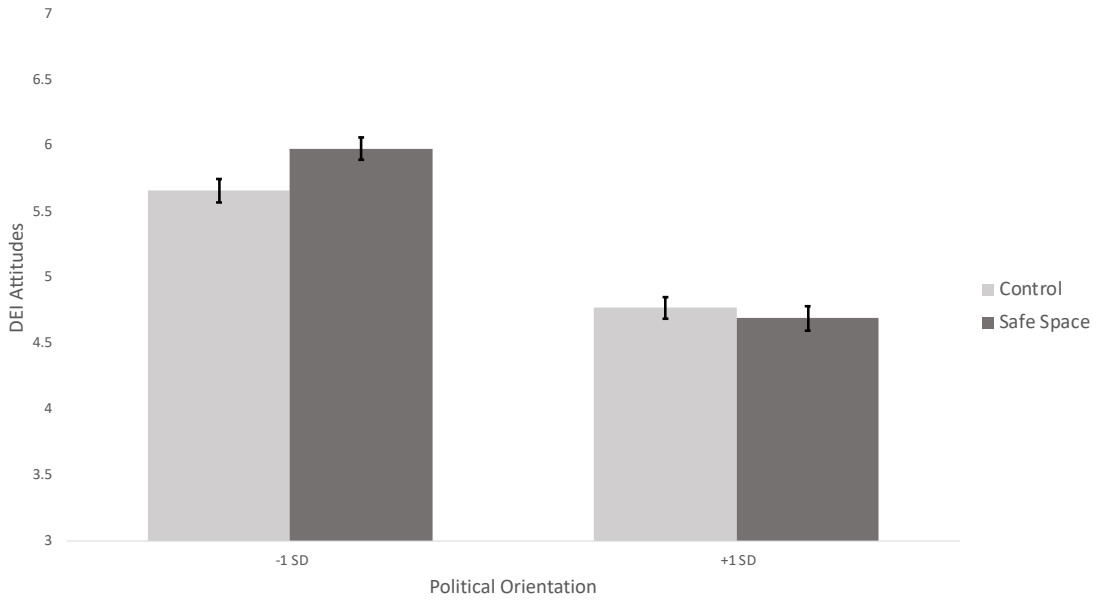

**Supplementary Fig 11.** Study 2 interaction between condition and experience with prejudice for prejudice expectations on campus. Error bars represent +/- 1 Standard Error.




**Supplementary Fig 12.** Study 2 interaction between condition and beliefs about Safe Spaces as protective (vs. coddling) for prejudice expectations in the US. Error bars represent +/- 1 Standard Error.




**Supplementary Fig 13.** Study 2 interaction between condition and political orientation for intentions to confronting prejudice. Error bars represent +/- 1 Standard Error.




**Supplementary Fig 14.** Study 2 interaction between condition and political orientation for attitudes toward diversity, equity, and inclusion. Error bars represent +/- 1 Standard Error.

## Study 3

### Vignette

Study 3 used a different vignette than the other studies to adapt the research question to a new context (the workplace) with a new institutional identity safety signal (an organizational commitment to diversity and inclusion). This vignette is below.

> It is 7:30 am on a Monday morning as John's alarm clock begins to ring incessantly. John is a recent college graduate, and like many of his peers, he's looking to start his career. Like any other day, he will work on various job applications, but today there will be a small addition to his list of things to do. For the first time in the past few weeks, John has landed an interview at a local marketing firm called Centium. He's scheduled to spend a couple hours there, as he learns about the company and interviews with a few different team leaders.
>
> John's first visit is with a representative from human resources. Arriving at the representative's office several minutes early, John takes a seat and begins to play a game on his phone while waiting patiently for his appointment. After a few minutes John becomes bored of his phone and begins to look over some materials that Centium had sent him, which he had printed out in advance.

First, John reads though the job description, which includes the day-to-day tasks and responsibilities that the job entails. Next, John reads though the opportunities for promotion, which includes details around financial bonuses and the career paths of others who have taken similar jobs at Centium. [Finally, John reads though the company's statement of support for diversity and inclusion policies, which says "Centium is an equal opportunity employer that celebrates diversity and is committed to creating an inclusive environment for all employees."] / [Finally, John reads though the company's statement of support for green and sustainable business practices, which says "Centium is an environmentally-friendly company that values sustainability and is committed to creating green business practices."]

**Measures**

Below we list Study 3 measures that were not reported or fully explicated in the main text.

**Values safety compared to other companies.** Participants answered, “To what extent do you think Centium values diversity and inclusion, as compared to other marketing companies?” from 1 (A lot less) to 7 (A lot more).

**Values environment compared to other companies.** Participants answered, “To what extent do you think Centium values green and sustainable business practices, as compared to other marketing companies?” from 1 (A lot less) to 7 (A lot more).

**Manipulation check.** Participants answered, “What (if any) things were included in the materials that Centium sent John? (Choose all that apply)” using the following options: Statement about diversity and inclusion, Job description, Opportunities for promotion, Statement about sustainability and environmentalism.

## Results

Supplementary Table 6 shows tests of how race, gender, and sexual orientation moderate the effect of condition in Study 3. Supplementary Table 7 shows tests of how continuous

moderators (e.g., political orientation) moderate the effect of condition in Study 3. We include figures that represent any moderation by the continuous variables.

Table 11

Study 3 Moderation by Race, Gender, and Sexual Orientation

| | Non-White | | White | | Interaction statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | M, Control | M, DEI Statement | M, Control | M, DEI Statement | df(error) | F | *p* | partial η2 |
| Prejudice in company | 29.67 [23.58, 35.76] | 31.54 [26.48. 36.60] | 24.87 [21.61, 28.12] | 28.70 [25.39, 32.00] | 538 | 0.2 | 0.67 | 0.000 |
| Prejudice in city | 40.53 [34.50, 46.57] | 38.21 [33.20, 43.23] | 34.01 [30.78, 37.24] | 40.29 [37.02, 43.57] | 538 | 3.5 | 0.06 | 0.006 |
| Prejudice in other companies | 41.40 [35.57, 47.22] | 43.63 [38.79, 48.47] | 34.55 [31.43, 37.66] | 39.83 [36.67, 42.99] | 538 | 0.5 | 0.50 | 0.001 |
| Prejudice in US | 56.88 [50.50, 63.26] | 57.71 [52.41, 63.02] | 47.64 [44.23, 51.05] | 52.26 [48.80, 55.73] | 538 | 0.6 | 0.44 | 0.001 |
| Confront Prejudice | 5.49 [5.19, 5.80] | 5.48 [5.23, 5.73] | 5.48 [5.31, 5.64] | 5.55 [5.38, 5.71] | 538 | 0.1 | 0.71 | 0.000 |
| DEI attitudes | 5.33 [5.02, 5.65] | 5.41 [5.15, 5.67] | 5.13 [4.96, 5.30] | 5.22 [5.05, 5.39] | 538 | 0.0 | 0.96 | 0.000 |
| Affirmative Action attitudes | 3.73 [3.48, 3.99] | 3.70 [3.49, 3.91] | 3.35 [3.21, 3.48] | 3.41 [3.28, 3.55] | 538 | 0.3 | 0.59 | 0.001 |
| | Women | | Men | | Interaction statistics | | | |
| | M, Control | M, DEI Statement | M, Control | M, DEI Statement | df(error) | F | *p* | partial η2 |
| Prejudice in company | 27.54 [23.67, 31.41] | 29.34 [25.38, 33.30] | 23.98 [19.73, 28.24] | 29.61 [25.71, 33.50] | 538 | 0.9 | 0.35 | 0.002 |
| Prejudice in city | 37.94 [34.10, 41.78] | 39.99 [36.06, 43.93] | 32.30 [28.08, 36.53] | 39.19 [35.23, 43.06] | 538 | 1.4 | 0.23 | 0.003 |
| Prejudice in other companies | 38.27 [34.56, 41.98] | 43.12 [39.32, 46.93] | 33.41 [29.33, 37.49] | 38.89 [35.16, 42.63] | 538 | 0.0 | 0.87 | 0.000 |
| Prejudice in US | 53.47 [49.42, 57.51] | 57.29 [53.15, 61.44] | 45.45 [41.01, 49.90] | 50.39 [46.32, 54.57] | 538 | 0.1 | 0.79 | 0.000 |
| Confront Prejudice | 5.68 [5.49, 5.87] | 5.59 [5.39, 5.79] | 5.19 [4.98, 5.40] | 5.45 [5.26, 5.64] | 538 | 3.0 | 0.08 | 0.006 |
| DEI attitudes | 5.39 [5.19, 5.59] | 5.38 [5.17, 5.58] | 4.88 [4.66, 5.10] | 5.17 [4.97, 5.37] | 538 | 2.1 | 0.15 | 0.004 |
| Affirmative Action attitudes | 3.65 [3.49, 3.81] | 3.63 [3.47, 3.80] | 3.20 [3.02, 3.37] | 3.36 [3.20, 3.52] | 538 | 1.1 | 0.29 | 0.002 |
| | Non-heterosexual | | Heterosexual | | Interaction statistics | | | |
| | M, Control | M, DEI Statement | M, Control | M, DEI Statement | df(error) | F | *p* | partial η2 |
| Prejudice in company | 36.44 [27.67, 45.22] | 43.57 [36.08, 51.06] | 24.72 [21.74, 27.70] | 27.24 [24.31, 30.17] | 536 | 0.5 | 0.46 | 0.001 |
| Prejudice in city | 46.59 [37.87, 55.32] | 51.49 [44.03, 58.94] | 34.05 [31.08, 37.01] | 37.98 [35.06, 40.89] | 536 | 0.0 | 0.88 | 0.000 |
| Prejudice in other companies | 46.82 [38.38, 55.26] | 53.08 [45.87, 60.29] | 34.84 [31.97, 37.71] | 39.11 [36.29, 41.93] | 536 | 0.1 | 0.74 | 0.000 |
| Prejudice in US | 59.70 [50.41, 69.00] | 64.14 [56.20, 72.07] | 48.43 [45.27, 51.58] | 52.64 [49.53, 55.74] | 536 | 0.0 | 0.97 | 0.000 |
| Confront Prejudice | 6.07 [5.63, 6.52] | 5.67 [5.29, 6.04] | 5.41 [5.26, 5.56] | 5.51 [5.36, 5.65] | 536 | 2.6 | 0.11 | 0.005 |
| DEI attitudes | 5.94 [5.48, 6.40] | 5.47 [5.08, 5.86] | 5.08 [4.93, 5.24] | 5.25 [5.10, 5.40] | 536 | 3.8 | 0.05 | 0.007 |
| Affirmative Action attitudes | 3.83 [3.46, 4.20] | 3.65 [3.33, 3.97] | 3.39 [3.26, 3.52] | 3.48 [3.35, 3.60] | 536 | 1.0 | 0.32 | 0.002 |

Table 12

Study 3 Moderation by Continuous Moderators

| | Interaction statistics - Political Orientation | | | |
|---|---|---|---|---|
| | df(error) | F | *p* | partial η2 |
| Prejudice in company | 541 | 5.8 | 0.02 | 0.011 |
| Prejudice in city | 541 | 4.3 | 0.04 | 0.008 |
| Prejudice in other companies | 541 | 1.6 | 0.21 | 0.003 |
| Prejudice in US | 541 | 1.0 | 0.31 | 0.002 |
| Confront Prejudice | 541 | 0.3 | 0.57 | 0.001 |
| DEI attitudes | 541 | 0.0 | 1.00 | 0.000 |
| Affirmative Action attitudes | 541 | 0.0 | 0.86 | 0.000 |
| | Interaction statistics - Prejudice Experience | | | |
| | df(error) | F | *p* | partial η2 |
| Prejudice in company | 540 | 2.0 | 0.16 | 0.004 |
| Prejudice in city | 540 | 0.0 | 0.98 | 0.000 |
| Prejudice in other companies | 540 | 0.0 | 0.85 | 0.000 |
| Prejudice in US | 540 | 0.6 | 0.44 | 0.001 |
| Confront Prejudice | 540 | 0.2 | 0.67 | 0.000 |
| DEI attitudes | 540 | 0.1 | 0.79 | 0.000 |
| Affirmative Action attitudes | 540 | 2.4 | 0.12 | 0.004 |

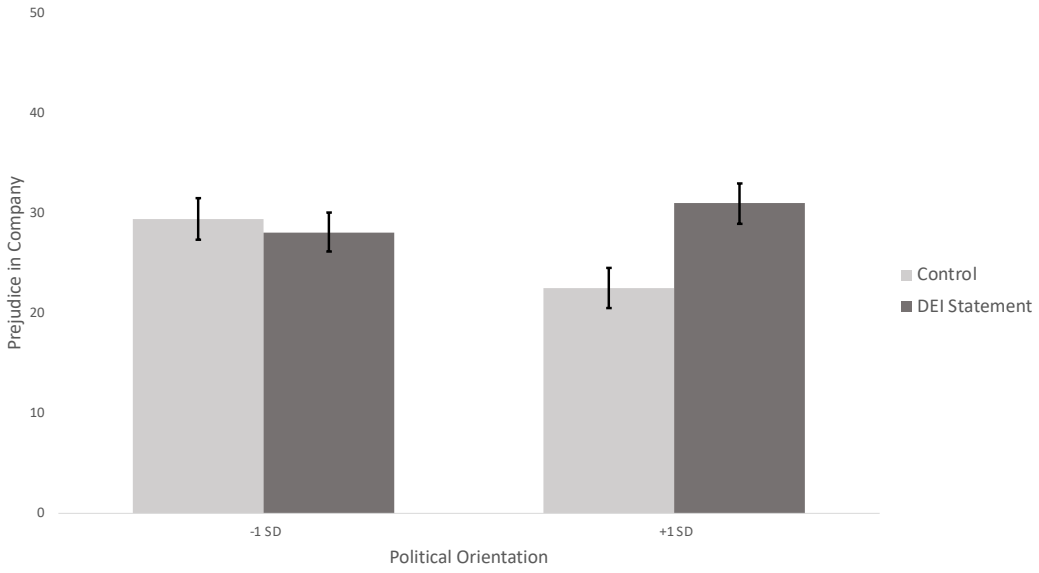



**Fig 15.** Study 3 interaction between condition and political orientation for prejudice expectations in the company. Error bars represent +/- 1 Standard Error.

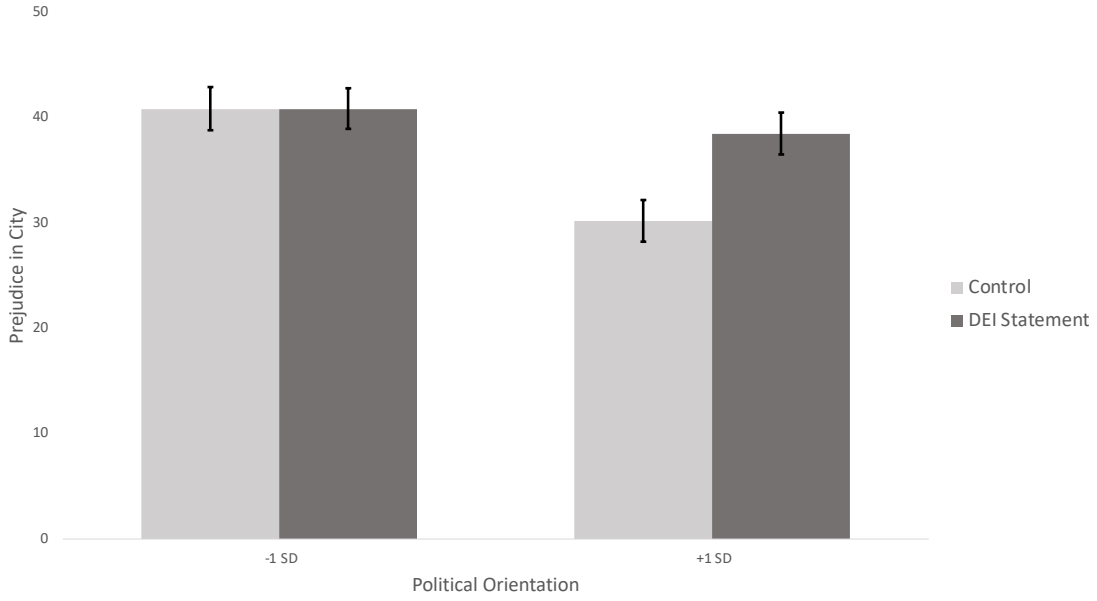



**Fig 16.** Study 3 interaction between condition and political orientation for prejudice expectations in the city. Error bars represent +/- 1 Standard Error.

## Study 4

### Method

#### Measures

Below we list Study 4 measures that were not reported or fully explicated in the main text.

**Problem is solved.** Participants rated their agreement with three items, "Given the sign in the advisor's office, the problem has been solved [in the office] / [on campus] / [in the United States]" from 1 (Strongly Disagree) to 7 (Strongly Agree).

**Evidence of caring.** Participants rated their agreement with three items, "The sign in the advisor's office signals that [the advisor] / [people on campus] / [people in the United States] care(s) a lot about the problem" from 1 (Strongly Disagree) to 7 (Strongly Agree).

## Results

Supplementary Table 8 shows tests of how continuous moderators (e.g., political orientation) moderate the effect of condition in Study 4. Supplementary Table 9 shows tests of how race, gender, and sexual orientation moderate the effect of condition in Study 4.

Table 13

Study 4 Moderation by Continuous Moderators

| | Interaction statistics - Protective Beliefs | | | |
|---|---|---|---|---|
| | df(error) | F | *p* | partial η2 |
| Prejudice in office | 786 | 1.7 | 0.20 | 0.002 |
| Prejudice on campus | 787 | 0.8 | 0.38 | 0.001 |
| Prejudice in US | 787 | 0.6 | 0.43 | 0.001 |
| Movements/Policies | 787 | 0.5 | 0.48 | 0.001 |
| SPLC Donations | 762 | 0.1 | 0.79 | 0.000 |

Table 14

Study 4 Moderation by Race, Gender, and Sexual Orientation

| | Non-White | | White | | Interaction statistics | | | |
|---|---|---|---|---|---|---|---|---|
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 25.29 [19.97, 30.62] | 25.21 [21.33, 29.10] | 15.60 [12.37, 18.84] | 16.00 [13.67, 18.33] | 807 | 0.0 | 0.90 | 0.000 |
| Prejudice on campus | 34.89 [29.15, 40.64] | 43.45 [39.27, 47.64] | 28.94 [25.46, 32.42] | 36.95 [34.44, 39.47] | 808 | 0.0 | 0.90 | 0.000 |
| Prejudice in US | 60.05 [54.15, 65.96] | 63.48 [59.17, 67.78] | 47.60 [44.02, 51.19] | 55.77 [53.19, 58.35] | 808 | 1.2 | 0.28 | 0.001 |
| Movements/Policies | 63.38 [57.36, 69.39] | 66.56 [62.17, 70.95] | 54.49 [50.84, 58.14] | 57.19 [54.56, 59.83] | 808 | 0.0 | 0.91 | 0.000 |
| SPLC Donations | .19 [.12, .25] | .26 [.21, .30] | .13 [.10, .17] | .12 [.09, .15] | 783 | 3.1 | 0.08 | 0.004 |
| | Women | | Men | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 20.21 [16.31, 24.11] | 19.11 [16.30, 21.92] | 15.31 [11.27, 19.36] | 17.89 [14.94, 20.85] | 802 | 1.1 | 0.30 | 0.001 |
| Prejudice on campus | 32.82 [28.72, 36.92] | 41.64 [38.67, 44.60] | 26.45 [22.19, 30.71] | 35.76 [32.65, 38.88] | 803 | 0.0 | 0.89 | 0.000 |
| Prejudice in US | 53.90 [49.68, 58.12] | 62.71 [59.65, 65.76] | 46.13 [41.74, 50.52] | 52.95 [49.75, 56.16] | 803 | 0.3 | 0.60 | 0.000 |
| Movements/Policies | 59.55 [55.21, 63.89] | 61.51 [58.37, 64.65] | 53.59 [49.07, 58.11] | 58.18 [54.88, 61.48] | 803 | 0.4 | 0.51 | 0.001 |
| SPLC Donations | .16 [.11, .20] | .15 [.12, .18] | .14 [.09, .18] | .16 [.13, .20] | 777 | 0.6 | 0.53 | 0.001 |
| | Non-heterosexual | | Heterosexual | | Interaction statistics | | | |
| | M, Control | M, Safe Space | M, Control | M, Safe Space | df(error) | F | *p* | partial η2 |
| Prejudice in office | 30.66 [22.16, 39.15] | 30.95 [25.99, 35.92] | 16.67 [13.77, 19.58] | 16.05 [13.88, 18.21] | 804 | 0.0 | 0.86 | 0.000 |
| Prejudice on campus | 46.93 [37.78, 56.09] | 44.99 [39.64, 50.34] | 28.25 [25.13, 31.34] | 37.44 [35.11, 39.77] | 805 | 3.7 | 0.05 | 0.005 |
| Prejudice in US | 68.90 [59.46, 78.34] | 65.72 [60.21, 71.23] | 48.54 [45.23, 51.68] | 56.32 [53.92, 58.73] | 805 | 3.5 | 0.06 | 0.004 |
| Movements/Policies | 67.71 [58.05, 77.37] | 68.69 [63.05, 74.33] | 55.44 [52.14, 58.73] | 58.01 [55.55, 60.47] | 805 | 0.1 | 0.79 | 0.000 |
| SPLC Donations | .23 [.12, .33] | .24 [.19, .30] | .14 [.11, .18] | .14 [.11, .17] | 780 | 0.1 | 0.76 | 0.000 |