frontiers | Frontiers in Social Psychology

TYPE Original Research
PUBLISHED 27 March 2024
DOI 10.3389/frsps.2024.1260974

OPEN ACCESS

EDITED BY
Hisham M. Abu-Rayya,
University of Haifa, Israel

REVIEWED BY
Allon Vishkin,
Technion Israel Institute of Technology, Israel
Brianne Hastie,
Murdoch University, Australia

*CORRESPONDENCE
Thomas Ian Vaughan-Johnston
✉ thomasvaughanjohnston@gmail.com

RECEIVED 18 July 2023
ACCEPTED 14 March 2024
PUBLISHED 27 March 2024

CITATION
Vaughan-Johnston TI, Nguyen A and Jacobson JA (2024) A surprising lack of consequences when constraining language. *Front. Soc. Psychol.* 2:1260974. doi: 10.3389/frsps.2024.1260974

A surprising lack of consequences when constraining language

Thomas Ian Vaughan-Johnston[1]*, Andrew Nguyen[2] and Jill A. Jacobson[3]

[1]Department of Psychology, Cardiff University, Cardiff, United Kingdom, [2]Department of Psychology, Mississippi University for Women, Columbus, IN, United States, [3]Department of Psychology, Queen's University, Kingston, ON, Canada

Introduction: Labels considered normatively appropriate for specific social identity groups change. Researchers have examined the effects of censorship and slur usage, but minimal research examines the psychological consequences of imposing new language constraints on people.

Methods: Across four samples of university students (N_{total} = 997), we sought participants' compliance in avoiding usage of numerous commonplace group labels while they wrote essays about obese people (Sample 1) or specific ethnic groups (Samples 2-4).

Results: We observed consistently high compliance rates: participants either invented novel terminology to describe the group or avoided group labels entirely. We observed a substantial *absence of* task discomfort, attitudinal shifts regarding the group, or motivational shifts, according to Bayesian analyses. Nor did we detect negative effects of language constraint among people who saw themselves as opposed to censorship.

Discussion: Although free speech and respectful language remain a multifaceted social debate, our findings show that university students are willing to follow even completely contrived language directives when describing social identity groups and to do so without substantial discomfort or backlash against those groups.

KEYWORDS

communication, compliance, censorship, free speech, null findings, unintended consequences

Introduction

Many hotly-debated social issues of the 21st century involve the use of labels for social identity groups. Some of these issues are moral questions: for instance, what labels should be used for specific social identity groups (e.g., "Native American," "Aboriginal," vs. "Indigenous"), and who is to decide this? Assuming normatively approved labels can be determined, however, many at least partially empirical questions are also raised by these debates. What messages or interventions can and should be used to encourage or pressure people into using these labels? Can the mass usage of alternative terms (e.g., "differently abled" vs. "disabled" or "handicapped") shift attitudes toward the group being labeled, or will revised terms provoke a "euphemism treadmill" by which old attitudes are simply transferred to the new terms (Greer, 1971; Pinker, 1994)? What compliance strategies are likely to stimulate resentment (e.g., provoking concerns about "language policing" and "political correctness"; e.g., Haidt, 2016) vs. being accepted without resistance? We think psychological science has given surprisingly little direct attention to these empirical issues, and in the present work we attempt to provide a "lightning in the bottle" demonstration of some relevant processes.

Choices about labels have consequences for members of threatened social identity groups. For instance, more negative attitudes toward gay people are activated in heterosexual people who are exposed to derogatory labels (i.e., slurs) vs. non-derogatory labels for gay people (Carnaghi and Maass, 2007). On the other hand, concerns are sometimes raised that efforts toward "political correctness" might result in backlash effects (i.e., negative attitudes toward the protected group, driven by reactance or avoidance), or the "chilling" of free speech (e.g., Strauts and Blanton, 2015; Haidt, 2016; Read, 2018). Concerns about political correctness tap into underlying political and moral concerns that obviously transcend the present investigation, which attempts to address some specific empirical questions. We tested whether soliciting compliance in avoiding certain group labels generates hostility or backlash effects among university students.

Multiculturalism and diversity initiatives

Social psychologists have developed increasingly sophisticated techniques to reduce negative attitudes held toward social identity groups (Allport, 1954; Hornsey and Hogg, 2000; Kawakami et al., 2007; Dovidio et al., 2008; Page-Gould et al., 2008; Johnson et al., 2018) or encourage multiculturalism more broadly (Rios and Wynn, 2016). A broad literature examines how diversity can be increased, and the benefits of multiculturalism (Crisp and Turner, 2011).

However, psychologists are beginning to also probe how diversity and multiculturalism initiatives can provoke resistance and backlash effects. Pushback toward prejudice-reduction and pro-diversity initiatives has often been observed (e.g., Vertovec and Wessendorf, 2010; Saad, 2020). Psychological interventions often provoke unwanted, unintended consequences (Wilson, 2011; Peters et al., 2014). Interventions designed to reduce gender bias may accidentally increase gender bias (Caleo and Heilman, 2019). People may become angry and show negative attitudinal shifts when pressured to engage in behaviors favorable toward minoritized groups (Plant and Devine, 2001). Interracial interactions, intended to improve intergroup attitudes through contact (e.g., Pettigrew and Tropp, 2006) may sometimes produce negative cognitive and emotional experiences toward the target or other prejudicial thinking (Shelton et al., 2005; Richeson and Shelton, 2007; Legault et al., 2011; Cooley et al., 2019). We think that attempts to control people's language could also prompt unintended negative reactions.

Why language control compliance may cause issues

An important social change relevant to academic/institutional settings and the broader public is the movement to have people comply with using specific terms for specific social identity groups (Marks, 2014; Indigenous Corporate Training Inc., 2016; National Assembly of State Arts Agencies, 2020; American Psychological Association, 2022). For instance, Canadians are asked to say "indigenous people," not "native Americans," "Indians," or "aboriginal people" (Indigenous Corporate Training Inc., 2016). Concerns about politically correct speech such as appropriate group labels has been a concern historically and more recently (for a review, see Henderson, 2003). However, researchers seldom consider the possible barriers involved in securing people's compliance in using (or avoiding) target group labels, and the little work in this area generally focuses on exposure to blatantly negative group labels (i.e., slurs; Carnaghi and Maass, 2007; Croom, 2011; Jeshion, 2013), rather than more innocuous terms.

Empirical evaluation of strategies to change language usage remains an unresolved social problem. One theory often invoked in the political correctness debate is psychological reactance (Brehm, 1966; Brehm and Brehm, 1981). According to reactance theory, "threat to or loss of a freedom motivates [an] individual to restore that freedom" (Brehm and Brehm, 1981, p. 4). Reactance can have clear relevance to anti-prejudice or language constraint interventions, which may threaten some people's feeling of freedom to think, speak, and act freely toward members of threatened social identity groups (also see Chen et al., 2015; Munger, 2017). Reactance to control attempts may manifest in a variety of ways, such as reactant people seeking to learn more about a banned topic (Worchel et al., 1975); and negative cognitions, affect, attitudes, or behavioral intentions toward the prescribed behaviors (Dillard and Shen, 2005). Freedom threats may even be conceptualized as threats to one's sense of self (Graupmann, 2018) or group identity (Kachanoff et al., 2022).

If securing people's compliance in using group labels produces reactance-related threats in those targets, we might also anticipate especially positive (negative) reactions from people who are relatively supportive (unsupportive) of censorship efforts that favor diversity/multiculturalism efforts. That is, language constraints will promote the preferences of people who support censorship as a social strategy to advance their social goals (e.g., Ashokkumar et al., 2020; Clark and Winegard, 2020; Costello et al., 2022). To that end, in the present work, we considered whether people who censoring language in the name of progressive values might have more positive reactions to our language control instructions.

When language control compliance may cause issues

Rather than being an invariant psychological response to any request for language compliance, people might only dislike interventions when they are accompanied by specific arguments or justifications. In university communities, for example, language compliance requests will often be accompanied by any guiding rationale ("do not use the word X, *because*…"). Anti-prejudice interventions may work most effectively when they focus on moral issues raised by prejudicial attitudes or behaviors, consistent with the "that's wrong" approach advocated by Johnson et al. (2018). Experimentally examining distinct framing techniques may produce interesting insights. For instance, people cannot abstractly judge which speech-acts they will consider offensive when they are actually exposed to them (Almagro et al., 2021). Similarly, which framing factors will shape people's willingness to abstain from using language (arbitrarily designated as) "offensive"

may not be intuitively obvious. We considered three types of framing consideration.

Positive vs. negative reasons

An example of a positive reason is that complying would showcase one's multicultural values. An example of a negative reason is that failure to comply could harm the target group's mental health because exposure to such language is harmful. In many domains, positive and negative information has asymmetrical effects (Baumeister et al., 2001; Rozin and Royzman, 2001; Fredrickson, 2013), so we wanted to consider the possibility of differential consequences.

Justification

Second, we varied the type of moral justification provided for language constraints, from consequentialist to deontological. *Consequentialist* morals appraise actions by considering what good or bad consequences arise from those actions (i.e., "speak this way because then something good will happen"), whereas *deontological* morals appraise actions' inherent qualities (i.e., "speak this way because it is inherently good to do so"). Consequentialist rationales often are used to persuade people to follow language directives (e.g., National Assembly of State Arts Agencies, 2020; American Psychological Association, 2022). However, consequentialist arguments can have downsides. For instance, people are less likely to be judged as moral when they justify actions through a consequentialist (vs. deontological) lens (Everett et al., 2018), and people should be more resistant to an intervention imposed by an immoral source. Therefore, deontological reasons may be more effective at securing compliance, or avoiding deleterious psychological consequences for compliers.

Arbitrary motivations

Finally, we considered that providing no justification at all might produce distinct effects from providing any justification. For instance, reactance concerns and persuasive backfire may be increased when task instructions suggest the experimenter's persuasive intent (Wicklund et al., 1970; Brauer et al., 2012), so ironically language constraint might operate most effectively without rhetorical justification. However, people are also more willing to comply when provided with even vacuous justifications from a requester (Langer et al., 1978), which might suggest that arbitrary requests may be especially resisted or disliked.

The present research

To investigate the above ideas, we exposed participants to a novel paradigm in which they were to avoid using a set of completely commonplace group labels when writing brief essays about those groups. We focused on university students because universities are often intellectual and legal battlefields for disputes about free speech, group labels, and prejudice (Byrne, 1990; O'Neil, 1997) and are often viewed descriptively or normatively as places where societal change may be initiated (Marullo and Edwards, 2000). Specifically, we examined the psychological consequences of prohibiting particular labels for specific social identity groups (i.e., *language constraints*). Our decision to prohibit (i.e., use a *proscriptive injunction*) rather than encourage (i.e., *prescriptive injunction*) specific language use is important because past work suggests that proscriptions generate more resistance and legitimacy concerns than do prescriptions (Pavey et al., 2022). Additionally, we chose to prohibit words presently in common usage (as opposed to words already considered inappropriate) because people tend to generate more resistance to novel restrictions to their freedom (which seem contestable) as opposed to established freedom restrictions (which seem uncontestable; Laurin et al., 2012). Thus, by maximizing situational factors that seem to generate resistance, the present work represents a strong test of the hypothesis that language guidelines generate negative responses. Most likely, our instructions will prompt immediate compliance; therefore, we tested both:

H1. A weak reactance hypothesis such that language constraints will increase reactance and decrease comfort.

H2. A strong reactance hypothesis that language constraints will produce more negative attitudes and behavioral intentions and decrease willingness to comply with subsequent language directives.

Still, we think that compliance directives may not have these effects for the reasons noted earlier. Because hypothesizing a null is counter to the null hypothesis significance testing (NHST) approach, we include Bayesian analyses to determine if we accumulated meaningful evidence for the null hypotheses (**H0**: Compliance instructions do not produce the effects denoted as H1/H2).

We also suspected that H1/H2 might depend on people's beliefs about diversity-related censorship activities, suggesting an interaction effect:

H3. Language constraints may lead to the negative consequences listed under H1 and H2 only for individuals low in pro-diversity censorship beliefs.

Additionally, we had more exploratory interests in the following questions:

Q1. Does positive vs. negative framing affect compliance rates or downstream consequences of compliance?

Q2. Does deontological vs. consequentialist framing affect compliance rates or downstream consequences of compliance?

Q3. Does framing in general (vs. providing only "arbitrary" or no specific justification) affect compliance rates or downstream consequences of compliance?

Method

Overview of the samples and integrated dataset

Our four experiments used very similar procedures and methods (see verbatim materials in SOM-1), integrating to $N = 997$ (see Table 1 for an overview of the samples). All samples were composed of Canadian university students, primarily white, primarily women (84% women, 15% men, 0.2% non-binary,

remainder PNA; M_{age} = 19.5, SD_{age} = 4.6; 75% White, 12% East Asian, 3% Black, 7% other, 3% PNA),[1] participating for course credit. In all cases, we attempted to get participants to comply with talking about a target social identity group while avoiding specific labels for that group (e.g., write an essay about Black targets while avoiding words like "Black" and "African-American").[2] The banned words were made up of words that would be commonly used in our university at the time of data collection. Given that some of our tests require such large sample sizes (e.g., Lakens et al., 2018), we decided to aggregate our data into an integrative data analysis (IDA; Curran and Hussong, 2009). The social group targeted for language constraints varied by sample (i.e., obese people in Sample 1, White or Black people in Samples 2–4),[3] as did what forms of language constraint condition were employed. Data/syntax are open at https://osf.io/vpm8a/.

Procedure

Phase 1: compliance request

Participants were initially introduced to the experiment's tasks: writing a few essays about a particular social group and answering some questionnaires. Before completing the writing tasks (Phase 2), they were randomly assigned to one of six between-participant conditions. In the *Control (No Constraint)* condition, no special instructions were given at this point. In all remaining *(Constraint)* conditions, however, participants were warned that they should "not use certain group labels when discussing the group…absolutely must avoid any slur language in this writing." For Sample 1 these words were "fat," "obese," "overweight," and "heavy." For Sample 2-4/s White Target conditions, these words were "white," "Europeans," and "Caucasians." For Sample 2-4/s Black Target conditions, these words were "black," "African-American," "African," "colored," and "person/people of color."

However, each Constraint condition differed in terms of how the constraint was justified. In the *Arbitrary Constraint* condition, no further justification was given. In the *Negative/Consequentialist* condition, we told participants that inappropriate language "makes people think that it is normal to dislike that group," that "groups exposed to such inappropriate language may feel socially isolated or rejected," and that "when individuals see this sort of inappropriate language they are more likely to experience anxious or depressive episodes." In the *Negative/Deontological* condition, we told participants that using "inappropriate group labels is simply the wrong thing to do," and that "it is problematic to be disrespectful, cruel, and indecent—it is intrinsically wrong." We characterized such language as "bigotry," and claimed that "bigotry and rejection of others are inherently bad things."

The remaining conditions modified the previous two conditions but using a positive rather than negative framing. In the *Positive/Consequentialist* condition, we told participants that using *appropriate* language generates positive attitudes toward the target group, that groups exposed to appropriate language feel socially welcomed and accepted, and that using appropriate language leads the speaker to have more positive attitudes toward that group. Finally, in the *Positive/Deontological* condition we stated that using appropriate language is "simply the right thing to do," that "it is an opportunity to be respectful, kind, and decent—it is intrinsically right," and that using appropriate language is an example of following "diversity," stating that, "diversity and acceptance of others are inherently good things."

Phase 2: writing tasks

Regardless of experimental condition, participants viewed a cluster of images showing four members of the target group, and reported what they would usually call people in that group using a textbox (following any rules imposed by Constraint conditions). Participants then wrote two paragraphs, each about the specific group they had just labeled. Our two writing prompts read as follows: "In this box, please write down your thoughts concerning your own personal experiences interacting with this [Sample 2–4: ethnic] group" (*Personal Interactions Task*), and "In this box, please write down your thoughts concerning how you think this [Sample 2-4: ethnic] group is treated in modern Canada" (*Cultural Context Task*). For each task, participants spent about 5 min writing. Thus, participants were being pressured into complying repeatedly across an extensive writing period.

Phase 3: reaction and moderator measures

Measures were filled out in the following order: (1) reactance emotions, (2) willingness for future compliance, (3) task comfort, (4) attitudes and behavioral intentions toward the target group in counterbalanced order, (5) motivations to control prejudice and censorship attitudes in counterbalanced order. Not all Samples included all measures, so degrees of freedom vary somewhat across tests. We summarize the measures briefly below, but see SOM-4 for more extensive descriptions.

Reactance emotions

We used Dillard and Shen's (2005) four item measure of reactance (*irritated*, *angry*, *annoyed*, and *aggravated*; rated 1 = *none of this feeling* to 11 = *a great deal of this feeling*; α =0.94, M = 4.09, SD = 1.50).

Future compliance

Participants rated how willing they would be to continue with the language directives in the future (Constraint conditions), or how willing they would be to follow language rules if we imposed

1 We did not collect demographics for all samples, but samples were drawn from the same university and demographics should therefore be very consistent. Any demographic information was collected at the end of the study. Ethnicity questions had fixed options which in some cases contradicted compliance instructions (e.g., "White / European") but note these were positioned immediately before debriefing.

2 According to pilot testing on 190 undergraduates drawn from the same population as the primary samples, on 9-point scales, African-Americans were seen as "suffering discrimination" (M = 6.2, SD = 2.0), as were overweight people (M = 5.4, SD = 2.1). White people were not seen as "suffering discrimination" (M = 2.1, SD = 1.2).

3 As we show in SOM-2, White vs. Black as a target group did not affect our results.

TABLE 1 Samples used to construct the integrative data analysis.

Sample #	Target stimuli	*N*	Conditions
1	Obese people	253	No Constraint Control (53); Negative/Consequentialist (49); Negative/Deontological (49); Positive/Consequentialist (52); Positive/Deontological (50).
2	White/Black people	357	No Constraint Control (76); Negative/Consequentialist (70); Negative/Deontological (64); Positive/Consequentialist (77); Positive/Deontological (70).
3	White/Black people	192	No Constraint Control (41); Negative/Consequentialist (41); Negative/Deontological (37); Positive/Consequentialist (35); Positive/Deontological (38).
4	White/Black people	195	No Constraint Control (66); Negative/Consequentialist (67); Arbitrary (62).

these rules on them (Control condition), from 1 (*very unlikely*) to 7 (*very likely*; $M = 4.95$, $SD = 1.68$).

Task comfort

Participants rated "How comfortable did these previous language-related tasks make you feel?" from 1 (*very uncomfortable*) to 7 (*very comfortable*; $M = 4.33$, $SD = 1.72$).

Attitudes

We averaged four items evaluating attitudes toward the target group (*deserving of social aid, intelligent, trustworthy, hardworking*; $\alpha = 0.76$; rated from 1 = *Strongly Disagree* to 7 = *Strongly Agree*; $M = 4.12$, $SD = 1.20$).

Behavioral intentions

We averaged four items evaluating positive behavioral intentions toward the target group (*playing sports, working on a project, having a conversation, playing a game*; $\alpha = 0.91$; rated from 1 = *Very Unlikely* to 7 = *Very Likely*; $M = 5.75$, $SD = 1.22$).

Motivations to control prejudice

We adapted Legault et al.'s (2007) 24-item scale by adding "at the moment" or "right now" to items to capture state motivations. Our factor analysis supported a five-factor solution (rather than Legault et al.'s six factors), corresponding to: *intrinsic motivations* ($\alpha = 0.87$; e.g., "Pleasure of being open-minded right now"; combining Legault et al.'s "intrinsic motivation" and "integrated regulation" subscales; $M = 4.93$, $SD = 1.31$), *identified regulation* ($\alpha = 0.84$; e.g., "Because right now I admire people who are egalitarian"; $M = 5.10$, $SD = 1.36$), *introjected regulation* ($\alpha = 0.82$; e.g., "Because I would feel guilty if I were prejudiced right now"; $M = 4.82$, $SD = 1.49$), *external regulation* ($\alpha = 0.81$; e.g., "Because I don't want people to think I'm narrow-minded at the moment"; $M = 3.35$ $SD = 1.43$), and *amotivation* ($\alpha = 0.83$; e.g., "I don't know, it's not very important to me right now"; $M = 2.47$, $SD = 1.31$).

Beliefs about censorship of anti-diversity

We adjusted 10 items from Hence and Wright's (1992) scale of censorship beliefs to capture pro-diversity censorship attitudes (e.g., "Intolerance (e.g., hate crimes) should not be depicted in television shows"); participants rated items from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). Only nine items loaded consistently ($\alpha = 0.82$) and were averaged ($M = 3.14$, $SD = 0.76$).

Debriefing

We debriefed participants and clarified that our specific language directives were contrived for the sake of the experiment. Because data were collected online, we do not have interview data with participants. However, the null results for reactance emotions and task comfort suggest that the paradigm was not particularly distressing for participants. Furthermore, no adverse events were reported for any of the studies.

Results

The following results are from the IDA which aggregates data from the four samples.

Compliance

Starting with the Control (no language constraint) conditions, 55% of Sample 1 participants referred to the targets using one or more of the labels that in the other conditions would be banned. The most frequent choice was "overweight." In contrast, in Sample 1, 2–6% (by condition) of Constraint condition participants used banned words. Constraint condition participants (told to avoid specific labels) were far less likely to use the banned words than Control participants (who were not told to avoid specific labels), $F_{(4, 247)} = 32.10$, $p < 0.001$, $\eta^2_p = 0.34$.

87%/88%/82% of Sample 2/3/4's Control (no language constraint) participants, respectively, referred to labels that in the Constraint conditions would be banned. In short, we succeeded in picking words that people conventionally use for these groups. In Samples 2/3/4, respectively, 27–33%/11–29%/13–34% (ranging by condition) of Constraint condition participants used banned words. Thus, constraint conditions greatly reduced those words' usages; Sample 2: $F_{(4, 352)} = 28.42$, $p < 0.001$, $\eta^2_p = 0.24$; Sample 3: $F_{(4, 187)} = 21.19$, $p < 0.001$, $\eta^2_p = 0.31$; Sample 4: $F_{(2, 192)} = 59.11$, $p < 0.001$, $\eta^2_p = 0.38$. These results support our prediction that compliance instructions would produce at least immediate behavior change. These findings also could be seen as a successful check for the compliance manipulation.

Compliant participants referred to White targets as "fair," "bright, light, happy," "English," "humans," "non-POC people," and other workarounds; and to Black targets as "Racialized people," "African," or "minority." Most participants used workarounds to avoid using banned terms, including "people" or "plus size," and interestingly some called the group "diverse" (presumably because

10.3389/frsps.2024.1260974

TABLE 2 Effects of specific constraint condition vs. control condition on all study outcomes.

	ANOVA omnibus test	Specific means (standard deviations) per experimental group						Bayes factor (H. favored)
		Control	Arbitrary constraint	Negative/ Conseq.	Negative/ Deon.	Positive/ Conseq.	Positive/ Deon.	
Compliance								
Compliance	$F_{(4, 755)} = 1.11$, $p = 0.351$, $\eta_p^2 = 0.006$	N/A	87.1% (33.8%)	80.6% (39.6%)	84.0% (36.8%)	78.7% (41.1%)	77.1% (42.2%)	75.1 (H_0)
Future compliance	$F_{(5, 990)} = 2.08$, $p = 0.066$, $\eta_p^2 = 0.010$	4.84 (1.84)	4.63 (1.67)	4.80 (1.69)	5.20 (1.59)	5.07 (1.63)	5.09(1.55)	21.1 (H_0)
Process variables								
Task comfort	$F_{(5, 990)} = 2.62$, $p = 0.023$, $\eta_p^2 = 0.013$	4.34 (1.80)	3.74 (1.67)	4.26 (1.66)	4.21 (1.65)	4.48 (1.75)	4.58 (1.71)	8.1 (H_0)
Reactance emotions	$F_{(4, 796)} = 0.60$, $p = 0.660$, $\eta_p^2 = 0.003$	4.19 (1.46)	N/A	4.10 (1.52)	3.94 (1.48)	4.06 (1.47)	4.13 (1.56)	223.6 (H_0)
Backlash effects toward target group								
Attitudes	$F_{(5, 987)} = 0.81$, $p = 0.546$, $\eta_p^2 = 0.004$	4.14 (1.28)	4.23 (1.05)	4.19 (1.22)	4.08 (1.22)	3.97 (1.17)	4.16 (1.13)	332.5 (H_0)
Behavioral intentions	$F_{(5, 988)} = 0.56$, $p = 0.733$, $\eta_p^2 = 0.003$	5.79 (1.21)	5.55 (1.35)	5.74 (1.28)	5.69 (1.23)	5.79 (1.16)	5.81 (1.16)	619.6 (H_0)
Beliefs about prejudice								
Intrinsic motivation	$F_{(4, 540)} = 0.10$, $p = 0.981$, $\eta_p^2 = 0.001$	4.88 (1.27)	N/A	4.99 (1.36)	4.92 (1.34)	4.92 (1.42)	4.95 (1.17)	281.0 (H_0)
Identified regulation	$F_{(4, 537)} = 68$, $p = 0.603$, $\eta_p^2 = 0.005$	5.08 (1.43)	N/A	5.07 (1.31)	5.15 (1.33)	4.97 (1.45)	5.26 (1.26)	101.5 (H_0)
Introject regulation	$F_{(4, 541)} = 0.48$, $p = 0.753$, $\eta_p^2 = 0.004$	4.91 (1.46)	N/A	4.92 (1.50)	4.70 (1.50)	4.73 (1.62)	4.82 (1.38)	147.3 (H_0)
External regulation	$F_{(4, 541)} = 0.91$, $p = 0.457$, $\eta_p^2 = 0.007$	3.50 (1.55)	N/A	3.45 (1.43)	3.29 (1.31)	3.18 (1.44)	3.31 (1.38)	68.6 (H_0)
Amotivation	$F_{(4, 540)} = 0.94$, $p = 0.439$, $\eta_p^2 = 0.007$	2.47 (1.28)	N/A	2.66 (1.35)	2.36 (1.24)	2.36 (1.36)	2.47 (1.32)	65.0 (H_0)
Censorship beliefs	$F_{(5, 986)} = 0.40$, $p = 0.852$, $\eta_p^2 = 0.002$	3.12 (0.84)	3.24 (0.87)	3.14 (0.74)	3.17 (0.69)	3.14 (0.70)	3.10 (0.77)	861.9 (H_0)

"Conseq." = consequentialist; "Deon." = deontological. Values refer to 95% confidence intervals. For the Bayes Factor column, H_0 = BF favors the null, H_1 = BF favors the alternative.

the pictures of obese individuals were racially diverse, and a mix of men/women). Many participants referred to "this ethnic group," "this ethnicity," "humans," and similar generic terms. We observed minimal reactance. A few Sample 1 participants used presumably facetious responses (e.g., "athletes"). A very small number of participants explicitly objected, for instance stating "white…I [sic] not offensive" in a White Target condition. In short, the Language Ban conditions did not make the task impossible for participants although language use often became vague or awkward, and most participants simply substituted alternative words.

Other effects (downstream consequences)

Effects of language constraint condition

For all remaining variables, we worked through a common set of analyses (full statistics reported in Tables 2, 3, respectively). First, we wanted to examine if our language compliance conditions caused negative reactions in participants. One-way ANOVA tests determined if any of the specific language constraint conditions produced unique effects compared to the rest (or vs. the Control condition). As Table 2 reveals, the effects for most variables were not significant; the only exception was task comfort, which we discuss briefly below. In sum, we did not find support for either the weak (H1) or strong (H2) reactance hypotheses.

Task comfort

We found significant evidence that task comfort differed across conditions. Because we did not have specific predictions about which particular cells might differ other than the Constraint conditions presumably reducing task comfort vs. Control, we used Bonferroni corrections to control for multiple testing when examining *post-hoc* comparisons between cells. The Arbitrary (no justification) condition only significantly differed from positive/deontological ($M_{diff} = -0.84$, $SE = 0.26$, $p = 0.017$). We are not inclined to interpret this effect any further, particularly because of the subsequent Bayesian analysis reported below.

TABLE 3 Effects of (any) constraint condition vs. control on study outcomes, moderated by pro-diversity censorship beliefs.

	Regression analysis	Bayes factor (BF)	BF favors
Compliance			
Compliance	Constraint B = 0.61 [0.53,0.70], $t = 13.84$, $p < 0.001$	>1,000 (Extreme)	Alternative
	Censorship Beliefs B =0.05 [0.00,0.10], $t = 2.05$, $p = 0.041$	3.5 (Moderate)	Null
	Interaction B = 0.03 [−0.08,0.14], $t = 0.53$, $p = 0.597$	9.2 (Moderate)	Null
Future compliance	Constraint B = −0.15 [−0.50,0.20], $t = -0.83$, $p = 0.405$	6.5 (Moderate)	Null
	Censorship Beliefs B = 0.31 [0.12,0.50], $t = 3.22$, $p = 0.001$	46.2 (Very Strong)	Alternative
	Interaction B =0.49 [0.04,0.94], $t = 2.13$, p =0.034	42.0 (Very Strong)	Alternative
Process variables			
Task comfort	Constraint B = −0.10 [−0.45,0.25], $t = -0.56$, $p = 0.577$	13.6 (Strong)	Null
	Censorship Beliefs B = −0.05 [−0.24,0.14], $t = -0.47$, $p = 0.639$	4.0 (Moderate)	Null
	Interaction B = −0.12 [−0.57,0.34], $t = -0.52$, $p = 0.607$	11.4 (Strong)	Null
Reactance emotions	Constraint B = 0.02 [−0.27,0.31], t = 0.12, $p = 0.902$	7.5 (Moderate)	Null
	Censorship Beliefs B = 0.08 [−0.08,0.24], $t = 0.99$, $p = 0.322$	3.5 (Moderate)	Null
	Interaction B = 0.17 [−0.21,0.54], $t = 0.88$, $p = 0.380$	12.4 (Strong)	Null
Backlash effects toward target group			
Attitudes	Constraint B = 0.03 [−0.25,0.30], t = 0.18, $p = 0.859$	14.1 (Strong)	Null
	Censorship Beliefs B = 0.22 [.08,0.37], $t = 2.97$, $p = 0.003$	193.2 (Extreme)	Alternative
	Interaction B = 0.34 [−0.01,0.69], $t = 1.89$, $p = 0.059$	7.7 (Moderate)	Null
Behavioral intentions	Constraint B = 0.04 [−0.21,0.29], t = 0.33, $p = 0.745$	11.9 (Strong)	Null
	Censorship Beliefs B = 0.25 [.11,0.38], $t = 3.55$, p <0.001	>1,000 (Extreme) 13.9	Alternative
	Interaction B = 0.08 [−0.25,0.40], t = 0.47, $p = 0.636$	(Strong)	Null
Beliefs about prejudice			
Intrinsic motivation	Constraint B = 0.07 [−0.19,0.34], t = 0.54, $p = 0.593$	9.6 (Moderate)	Null
	Censorship Beliefs B = 0.32 [.18,0.47], $t = 4.42$, p <0.001	699.3 (Extreme)	Alternative
	Interaction B = −0.01 [−0.35,0.34], $t = -0.04$, $p = 0.969$	10.5 (Strong)	Null
Identified regulation	Constraint B = 0.03 [−0.25,0.31], t = 0.22, $p = 0.829$	10.3 (Strong)	Null
	Censorship Beliefs B = 0.20 [.05,0.35], $t = 2.56$, $p = 0.011$	2.4 (Anecdotal)	Alternative
	Interaction B = 0.23 [−0.14,0.59], $t = 1.23$, $p = 0.219$	4.7 (Moderate)	Null
Introjected regulation	Constraint B = −0.11 [−0.42,0.20], $t = -0.69$, $p = 0.492$	8.0 (Moderate)	Null
	Censorship Beliefs B = 0.11 [−0.06,0.28], $t = 1.30$ $p = 0.193$	4.6 (Moderate)	Null
	Interaction B = 0.31 [−0.09,0.71], $t = 1.55$, $p = 0.122$	3.2 (Moderate)	Null
External regulation	Constraint B = −0.19 [−0.48,0.10], $t = -1.27$, $p = 0.206$	4.9 (Moderate)	Null
	Censorship Beliefs B = 0.09 [−0.07,0.24], $t = 1.05$, $p = 0.295$	5.8 (Moderate)	Null
	Interaction B = 0.30 [−0.08,0.68], $t = 1.57$, $p = 0.117$	3.7 (Moderate)	Null
Amotivation	Constraint B = 0.00 [−0.27,0.27], $t = -0.01$, $p = 0.990$	10.5 (Strong)	Null
	Censorship Beliefs B = −0.14 [−0.29,0.01], $t = -1.87$, $p = 0.062$	2.0 (Anecdotal)	Null
	Interaction B = −0.06 [−0.41,0.29], $t = -0.33$, $p = 0.739$	9.4 (Moderate)	Null

BF, Bayes Factor. Values in square brackets refer to 95% confidence intervals.

Bayesian analysis

Despite the results reported above, finding null effects in NHST does not provide clear support for the null hypothesis (see Wagenmakers et al., 2011). Furthermore, we were unclear how seriously to take the single significant effect on task comfort. Accordingly, we checked if we had accrued meaningful support for a null hypothesis. Therefore, we performed a Bayesian analysis focused on determining the Bayes Factor associated with support for the null over the alternative hypothesis using the *anovaBF* function from the BayesFactor package (Morey and Rouder, 2022) in R (R Core Team, 2022). We used the package's prior probability defaults because these defaults were specifically developed to be "general, broadly applicable" (Rouder et al., 2012, p. 356). Conventional standards suggest that Bayes Factors between 1 and 3 provide "anecdotal" support for a hypothesis, 3–10 provide "moderate" support, 10–30 offer "strong" support, 30–100 "very strong," and >100 "extreme." In Table 2, all BFs are expressed as BF_{01}, meaning that they express how many times more likely the data are under the null over the alternative hypothesis (i.e., larger numbers indicate greater support for the null hypothesis). In sum, the preponderance of evidence moderately to greatly favored the null hypothesis, H0. Additionally, the various framing conditions (i.e., Q1–Q3) did not make a substantive difference.

Bayesian analyses supported the null hypothesis, usually strongly.[4] In total, one Bayes Factor was in each of the "moderate" and "strong" evidence ranges, three in the "very strong," and seven in the "extreme" range. These analyses increase our confidence that despite creating substantial compliance among participants (at least two-thirds of participants always complied when directed, as discussed previously), our compliance request did not make them uncomfortable or feel reactance, did not lead to attitude or behavior-intentional backlash effects against the group protected by the language compliance instructions, did not shift people's motivations to control prejudice to a more external basis, and did not alter people's beliefs about the (lack of) value in using censorship to protect vulnerable social identity groups.

Overall effects of constraining language

Perhaps by including so many conceptually diverse language constraint conditions, we overlooked a specific contrast of interest: whether imposing language constraint instructions *at all* had aversive effects on people compared to the control condition which imposed no such rules. In Supplementary material (SOM-3), we test this possibility by comparing the Control group against an aggregation of all the Compliance groups, examining this contrast with independent-samples *t*-tests, Bayesian *t*-tests, and equivalence tests (Lakens, 2017; Lakens et al., 2018). Briefly, all significance tests were non-significant, and Bayesian tests produced substantial evidence favoring the null for all variables. Furthermore, equivalence tests run against two benchmarks (and a theory-derived effect size of d =0.41; the so-called "small" effect size of d = 0.20, Cohen, 1988, also see Richard et al., 2003) provided us with statistically significant basis to say that if constraining language had effects on any of our diverse outcomes, these effects may be ignored as unimportantly small by two distinct standards.

Moderation by censorship beliefs

Finally, we wondered if our usage of university students may have steered our results toward placidity. Universities including the one at which we collected data promote diversity, and students might therefore find even strongly-worded and unusual requests to avoid certain language choices acceptable and normal. We therefore tested the possibility that despite a general absence of concerning effects of language constraints, effects might at least emerge among the subset of participants least in favor of censorship in the name of social identity goals.[5] We, therefore, tested if any effects of language constraint (vs. control) were moderated by censorship beliefs (H3). We analyzed this research question using standard OLS regression but also the Bayes Factor associated with each parameter (contrast-coded main effect of constraint, centered main effect of censorship beliefs, and their interaction) tested against an intercept-only model. All results are reported in Table 3. The BF is always reported as >1 to ease comparison, so we note whether the BF supports the null (i.e., was BF_{01}) or the alternative hypothesis (i.e., was BF_{10}).

For 10 of 11 interaction terms, we found a non-significant effect and moderate or greater support for the null over alternative hypothesis using Bayesian analysis. Thus, even participants who saw themselves as relatively unsupportive of censorship to benefit progressive goals were indifferent to our language constraint intervention. In sum, most variables did not support H3.

In the case of future compliance, however, we detected a significant interaction effect, also corroborated by "very strong" support for the alternative over null hypothesis in Bayesian testing. This interaction is also tracked in Figure 1. As the figure illustrates, the effect of experimental condition (constraint conditions in blue, control condition in red) shifted based on participants' pro-diversity censorship beliefs. Participants lower in pro-diversity censorship (left side of figure) anticipated less compliance after undergoing a constraint manipulation whereas participants higher in pro-diversity censorship (right side of figure) anticipated significantly greater compliance after undergoing a constraint manipulation. A Johnson-Neyman analysis indicated that Constraint manipulations (vs. Control) prompted significantly less anticipated compliance among the 14% of participants most anti-censorship, and significantly more anticipated compliance among the 42% of participants most favoring pro-diversity censorship.

Furthermore, we found that people more supportive of censorship in the name of social justice (i) were also more willing

4 Of course, this contradicts the NHST testing in the case of comfort because Bayesian testing supported the null and NHST supported the alternative hypothesis. Data that reaches statistical significance in NHST while also supporting the null over alternative hypothesis are not necessarily surprising; for example, see Wagenmakers et al. (2011). It is most easily explained by observing that the effect was very small and detected because of our large sample size.

5 One may wonder if censorship beliefs truly were a moderator rather than a consequence of the compliance manipulation. Because Hence and Wright found a high (r = 0.83) three-week test-retest reliability of their scale, we assumed these beliefs have substantial trait-like variance and would probably not change based on our manipulation. As we earlier showed (Table 2), the manipulation indeed did *not* change these beliefs.

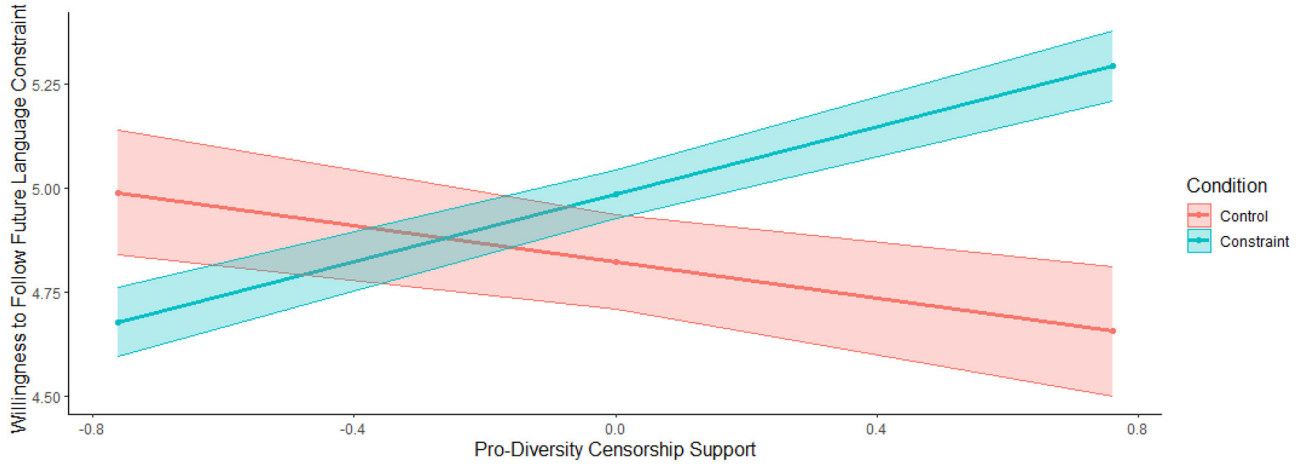

FIGURE 1
Willingness to follow future language constraints is influenced by language constraint condition and pro-diversity censorship support.

to comply with future language control instructions, (ii) had more positive opinions of the target groups, (iii) had more positive behavioral intentions toward the target groups, (iv) were more intrinsically motivated to control for their prejudices, and (v) had more identified-regulation motivation to control prejudice. A final effect suggested that censorship beliefs might be favorably related to (vi) avoidance of the target words, but the Bayesian analysis suggested this effect was so weak as to be more consistent with the null than the alternative hypothesis. These relationships are generally in line with our expectations, and help to establish that our novel measure of censorship attitudes showed sensible patterns of validity. That is, someone who has positive views of promoting diversity with authoritarian means would be likely to be personally okay with complying with new language rules, report favorable attitudes and behavioral intentions social groups identified as needing such protection, and have more positive and morally based desires to deal with prejudice.

General discussion

In summary, our four experiments covered diverse variables, targets, analyses, and message types. However, the key finding is straightforward: university students willingly followed arbitrary and frustrating language directives simply because we told them to. Participants readily adopted our new language conventions even when we gave no rationale whatsoever. There were no negative emotional or attitudinal shifts, problematic motivational styles, or adverse consequences observed across nearly a thousand participants in various analyses (Frequentist and Bayesian). People's beliefs about censorship had surprisingly little impact, except for those supporting censorship to promote diversity, who showed increased willingness to follow our future directives. Thus, participants complied with the act of abandoning frequently used words, effectively capturing the phenomenon of novel changes to group labels commonly seen in modern society.

Implications

On one hand the present results might be considered concerning. The vast majority of undergraduates obediently followed nonsensical instructions to avoid evaluatively-neutral words without resistance. Avoiding terms like "white" or "Caucasian" because we arbitrarily banned them as "offensive" might be seen as problematic. Students extreme malleability could be seen as a lack of critical discernment regarding reasonable vs. unreasonable language requests.

On the positive side, our data suggest real-world language constraints need not always lead to psychological issues among university students, even for those opposed to censorship. We intentionally created a strong situation to provoke backlash, including proscribing language, banning conventional words, and demanding compliance toward a group perceived as not needing support. Despite this setup, we observed minimal problematic reactions. Therefore, it's unlikely that real-world interventions, which offer alternatives, target disfavored language, and protect minoritized groups, would cause issues.

Obviously, a large literature on reactance and autonomy threats supports that people are often resistant or overtly hostile to attempts made to change their speech, attitudes, and behaviors (Brehm, 1966; Worchel et al., 1975; Brehm and Brehm, 1981; Dillard and Shen, 2005; Chen et al., 2015; Munger, 2017; Graupmann, 2018). Therefore, it is worth asking what considerations of the present work led to conditions in which people placidly tolerated a novel (and arguably absurd) demand to change their language.

One consideration is that university students are very frequently exposed to novel compliance requests about appropriate language usage (Roberts, 2017; Macnamara, 2022; for example news stories, see Anderson, 2022; Price, 2023), and are sometimes even paid to confront inappropriate language on campus (Coughlan, 2020). Thus, compliance in this domain may be a well-formed habit for students. When people become accustomed to a social norm involving a behavior such as language change, they

may become highly open to additional revisions in "approved terminology." In essence, having grown accustomed to one specific form of compliance, subsequent compliance requests may be highly successful (i.e., the foot-in-the-door technique; Freedman and Fraser, 1966; Burger, 1999; Pettigrew and Tropp, 2006), a phenomenon that our paradigm perhaps exploited. Past scholars have suggested that foot-in-the-door may work because consistent compliance is motivated by self-enhancement: one's past behaviors, being one's own, are perceived as good, making the present (similar) behavior also seem good (Cialdini and Goldstein, 2004). This may help to explain the lack of any negative psychological reactions from participants' (continued) compliance. However, since past compliance research seldom assesses psychological reactions directly, our work contributes to this area by measuring whether compliant participants felt any psychological resistance.

Another consideration is that people have strong desires not to *be* bigoted, and not to *seem* bigoted (e.g., Devine et al., 2002; Legault et al., 2007; against racial minorities). Compliance in our paradigm (i.e., avoiding words that we stated were offensive) might be perceived as consistent with either motivation. That is, whether a given participant was primarily motivated to merely avoid seeming bigoted, or whether they actually wished not to be bigoted, compliance was presumably a safer choice than resistance.

Limitations and future directions

Stakes and framing issues

Our paradigm can be considered low-stakes for participants in some respects. That is, they did not have to interact with other people while following language directives, and they were free to disregard the instructions outside of the experiment (which we made clear in the debriefing but would have been true regardless). We cannot entirely dismiss that they complied because they simply did not care very much about the task, but we do have a few counterpoints. First, the interaction of censorship beliefs by constraint on participants' intentions to follow future language directives suggests that the intervention was psychologically real enough that people's core values around speech and censorship polarized people's response in terms of behavioral intentions to comply later. Behavioral intentions, such as our willingness-to-comply measure, often predict behavior reasonably well (Webb and Sheeran, 2006), so we consider this finding to be noteworthy. Second, even if participants' compliance represented a superficial normative conformity to instructions (i.e., "I'll comply to not make waves"), normative influences often provoke changes in internal construal (Griffin and Buehler, 1993), so the high compliance rates might nonetheless be consequential.

One possibility is that our framing manipulations were ineffective because of their particular wording choices. The manipulations do have ecological validity in that they were directly modeled on real-world directives from companies (e.g., Indigenous Corporate Training Inc., 2016) and relevant organizations (e.g., media reference guides from GLAAD; e.g., GLAAD, 2024) that often use positive/negative and consequentialist/deontological framing and/or arbitrary justifications when arguing for appropriate speech. Thus, they validly represent the sort of messages commonly distributed to the public and capture the spirit of how such messages are circulated. Furthermore, often the specific wording of justifications is not what matters in compliance paradigms, but the mere provision of "any" reason (Langer et al., 1978). Nonetheless, more carefully optimized versions of our framings might produce different results. One interesting difference between our stimuli and the raw material is that frequently these sources refer to multiple reasons to follow language constraints. Future research could examine if more compliance or downstream consequences differ when multiple justifications are combined in the same intervention.

Might others resist more (or even less)?

Our sampling was narrowly focused on university students in a Western context. Future research might tackle this limitation by examining a few types of heterogeneity. First, past research suggests that cultures vary in the extent to which they cultivate a need to follow one's preferences. Savani et al. (2008) found that the association between preferences and choices was very pronounced among North Americans, but was attenuated among Indians. Assuming that people's conventional labels for groups can be considered a personal preference, examination of non-Western cultures that privilege personal preferences less could lead to higher compliance rates when securing agreement to proposed language changes (or given our very high overall compliance, greater compliance to more strongly-worded or invasive forms of the intervention).

Second, we examined beliefs about pro-diversity censorship because we wanted an individual difference moderator that was maximally likely to alter reactions to our language constraints. However, future work should draw samples that include a broader political spectrum including political conservatives, who often lodge objections against politically correct speech and might therefore react more negatively to language constraints (e.g., Fish, 1994; Wilson, 2020). Relatedly, our sample was primarily young adults who more often are strongly politically left or progressive (Electoral Calculus, 2019). Thus, our effects might have been stronger than if we had sampled older adults, given that progressive people might be more sympathetic to the ostensible intentions of our language compliance paradigm. Indeed, Proulx et al. (2022) showed that "mandated diversity" (which could include language constraints) is one of the beliefs that distinguishes these two left-wing groups (i.e., traditional liberals and political progressives).

Third, our samples were also mostly women. Laboratory research sometimes finds higher compliance rates from women vs. men (e.g., Tom and Granié, 2011) across miscellaneous domains, but such differences are usually modest (Eagly, 1983; Grosch and Rau, 2016). So gender probably does not account for our findings or limit their generalizability. Still, replicating our studies with more representative proportions of older conservative men and especially in a non-university context would be expected to change our baseline compliance rates.

Collective autonomy threats

In our experiments, participants were isolated individuals exposed to language compliance requests. However, one can

also imagine contexts in which a group of people might be exposed to a compliance request to use or avoid specific group labels. Interestingly, restricting groups' ability to communicate can result in decreased wellbeing for the people experiencing such constraints (Kachanoff et al., 2019, 2020, 2022). Therefore, shifting our research to a context of group discussion, in which the compliance request might be seen as an imposition on the group's autonomy, might generate collective autonomy threats and therefore stimulate more negative reactions. Researchers can thus continue to explore what determines whether people will comply with vs. resist language directives.

Conclusion

Public debates on free speech, language constraints, political correctness, etc., surpass the scope of a single research article. We aim to test claims from popular and academic sources on these matters. Our goal is not to take a stance on language constraints. The current findings represent an initial investigation into language control's downstream consequences. We hope these results spark more interest and provide evidence-based insights into heated social questions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving humans were approved by Queen's University General Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TV-J: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing. AN: Conceptualization, Methodology, Writing—review and editing. JJ: Formal analysis, Methodology, Project administration, Resources, Supervision, Validation, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frsps.2024.1260974/full#supplementary-material

References

Allport, G. W. (1954). *The Nature of Prejudice*. Boston: Addison-Wesley.

Almagro, M., Hannikainen, I. R., and Villanueva, N. (2021). Whose words hurt? Contextual determinants of offensive speech. *Pers. Soc. Psychol. Bull.* 48, 937–953. doi: 10.1177/01461672211026128

American Psychological Association (2022). *Inclusive Language Guidelines*. Available online at: https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines#

Anderson, N. (2022). *Linguists slam Cambridge University for Teaching 'Woke' German: Fury as Students are Taught 'Gender-Neutral' Version of the Language That Avoids Using Masculine Forms that are not 'Inclusive' to Non-Binary People*. Available online at: https://www.dailymail.co.uk/news/article-11348533/Linguists-slam-Cambridge-University-teaching-woke-version-gender-neutral-German.html

Ashokkumar, A., Talaifar, S., Fraser, W. T., Landabur, R., Buhrmester, M., Gómez, Á., et al. (2020). Censoring political opposition online: who does it and why. *J. Exp. Soc. Psychol.* 91, 104031. doi: 10.1016/j.jesp.2020.104031

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Rev. General Psychol.* 5, 323–370 doi: 10.1037/1089-2680.5.4.323

Brauer, M., Er-Rafiy, A., Kawakami, K., and Phills, C. E. (2012). Describing a group in positive terms reduces prejudice less effectively than describing it in positive and negative terms. *J. Exp. Soc. Psychol.* 48, 757–761. doi: 10.1016/j.jesp.2011.11.002

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. Academic Press.

Brehm, S. S., and Brehm, J. W. (1981). *Psychological Reactance: A Theory of Freedom and Control*. Academic Press.

Burger, J. M. (1999). The foot-in-the-door compliance procedure: a multiple-process analysis and review. *Pers. Soc. Psychol. Rev.* 3, 303–325. doi: 10.1207/s15327957pspr0304_2

Byrne, J. P. (1990). Racial insults and free speech within the university. *Georgetown Law J.* 79, 399.

Caleo, S., and Heilman, M. E. (2019). What could go wrong? Some unintended consequences of gender bias interventions. *Arch. Scient. Psychol.* 7, 71. doi: 10.1037/arc0000063

Carnaghi, A., and Maass, A. (2007). In-group and out-group perspectives in the use of derogatory group labels: gay vs fag. *J. Lang. Soc. Psychol.* 26, 142–156. doi: 10.1177/0261927X07300077

Chen, B., Van Assche, J., Vansteenkiste, M., Soenens, B., and Beyers, W. (2015). Does psychological need satisfaction matter when environmental or financial safety are at risk? *J. Happiness Stud.* 16, 745–766. doi: 10.1007/s10902-014-9532-5

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Clark, C. J., and Winegard, B. M. (2020). Tribalism in war and peace: the nature and evolution of ideological epistemology and its significance for modern social science. *Psychol. Inq.* 31, 1–22. doi: 10.1080/1047840X.2020.1721233

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Cambridge: Academic Press.

Cooley, E., Brown-Iannuzzi, J. L., Lei, R. F., and Cipolli, I. I. I., W. (2019). Complex intersections of race and class: Among social liberals, learning about White privilege reduces sympathy, increases blame, and decreases external attributions for White people struggling with poverty. *J. Exp. Psychol.: General* 148, 2218. doi: 10.1037/xge0000605

Costello, T. H., Bowes, S. M., Stevens, S. T., Waldman, I. D., Tasimi, A., and Lilienfeld, S. O. (2022). Clarifying the structure and nature of left-wing authoritarianism. *J. Pers. Soc. Psychol.* 122, 135. doi: 10.1037/pspp0000341

Coughlan, S. (2020). *Sheffield Students Paid to Tackle Racist Language on Campus*. Available online at: https://www.bbc.co.uk/news/education-51098539

Crisp, R. J., and Turner, R. N. (2011). Cognitive adaptation to the experience of social and cultural diversity. *Psychol. Bull.* 137, 242. doi: 10.1037/a0021840

Croom, A. M. (2011). Slurs. *Lang. Sci.* 33, 343–358. doi: 10.1016/j.langsci.2010.11.005

Curran, P. J., and Hussong, A. M. (2009). Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol. Methods* 14, 81. doi: 10.1037/a0015914

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., and Vance, S. L. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *J. Pers. Soc. Psychol.* 82, 835. doi: 10.1037/0022-3514.82.5.835

Dillard, J. P., and Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Commun. Monogr.* 72, 144–168. doi: 10.1080/03637750500111815

Dovidio, J. F., Glick, P., and Rudman, L. A. (2008). *On the Nature of Prejudice: Fifty Years After Allport*. Hoboken: John Wiley and Sons.

Eagly, A. H. (1983). Gender and social influence: a social psychological analysis. *Am. Psychol.* 38, 971. doi: 10.1037/0003-066X.38.9.971

Electoral Calculus (2019). *Three-D Politics and the Seven Tribes*. Available online at: https://www.electoralcalculus.co.uk/pol3d_main.html

Everett, J. A., Faber, N. S., Savulescu, J., and Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *J. Exp. Soc. Psychol.* 79, 200–216. doi: 10.1016/j.jesp.2018.07.004

Fish, S. (1994). *There's No Such Thing as Free Speech: And it's a Good Thing, Too*. Oxford: Oxford University Press.

Fredrickson, B. L. (2013). Positive emotions broaden and build. *Adv. Exp. Soc. Psychol.* 47, 1–53. doi: 10.1016/B978-0-12-407236-7.00001-2

Freedman, J. L., and Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *J. Pers. Soc. Psychol.* 4, 195. doi: 10.1037/h0023552

GLAAD (2024). *GLAAD Media Reference Guide*. Available online at: https://glaad.org/reference/communities-of-color/

Graupmann, V. (2018). Show me what threatens you, and I can tell who you are: perception of threat and the self. *Self Identity* 17, 407–417. doi: 10.1080/15298868.2017.1412346

Greer, G. (1971). *The Female Eunuch*. New York: Farrer, Straus and Giroux.

Griffin, D., and Buehler, R. (1993). Role of construal processes in conformity and dissent. *J. Pers. Soc. Psychol.* 65, 657. doi: 10.1037/0022-3514.65.4.657

Grosch, K., and Rau, H. A. (2016). "Gender differences in compliance: the role of social value orientation," in *GlobalFood Discussion Paper 88, University of Göttingen*. Available online at: https://www.econstor.eu/handle/10419/146902 (accessed Feb 22, 2024).

Haidt, J. (2016). Why concepts creep to the left. *Psychol. Inq.* 27, 40–45. doi: 10.1080/1047840X.2016.1115713

Henderson, A. (2003). What's in a Slur?. *Am. Speech* 78, 52–74. doi: 10.1215/00031283-78-1-52

Hornsey, M. J., and Hogg, M. A. (2000). Subgroup relations: a comparison of mutual intergroup differentiation and common ingroup identity models of prejudice reduction. *Pers. Soc. Psychol. Bull.* 26, 242–256. doi: 10.1177/0146167200264010

Indigenous Corporate Training Inc. (2016). *Indigenous Peoples Terminology Guidelines for Usage*. Available online at: https://www.ictinc.ca/blog/indigenous-peoples-terminology-guidelines-for-usage

Jeshion, R. (2013). Slurs and stereotypes. *Analytic Philosophy* 54, 314–329. doi: 10.1111/phib.12021

Johnson, I. R., Kopp, B. M., and Petty, R. E. (2018). Just say no!(and mean it): Meaningful negation as a tool to modify automatic racial attitudes. *Group Proc. Intergroup Relat.* 21, 88–110. doi: 10.1177/1368430216647189

Kachanoff, F. J., Gray, K., Koestner, R., Kteily, N., and Wohl, M. J. (2022). Collective autonomy: why groups fight for power and status. *Soc. Personal. Psychol. Compass* 16, e12652. doi: 10.1111/spc3.12652

Kachanoff, F. J., Kteily, N. S., Khullar, T. H., Park, H. J., and Taylor, D. M. (2020). Determining our destiny: Do restrictions to collective autonomy fuel collective action?. *J. Pers. Soc. Psychol.* 119, 600–632. doi: 10.1037/pspi0000217

Kachanoff, F. J., Taylor, D. M., Caouette, J., Khullar, T. H., and Wohl, M. J. (2019). The chains on all my people are the chains on me: restrictions to collective autonomy undermine the personal autonomy and psychological well-being of group members. *J. Pers. Soc. Psychol.* 116, 141–165. doi: 10.1037/pspp0000177

Kawakami, K., Phills, C. E., Steele, J. R., and Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: the impact of approach orientations on attitudes toward Blacks. *J. Pers. Soc. Psychol.* 92, 957–971. doi: 10.1037/0022-3514.92.6.957

Lakens, D. (2017). Equivalence tests: a practical primer for t-tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* 1, 1–8. doi: 10.1177/1948550617697177

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci.* 1, 259–269. doi: 10.1177/2515245918770963

Langer, E. J., Blank, A., and Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: the role of "placebic" information in interpersonal interaction. *J. Pers. Soc. Psychol.* 36, 635–642. doi: 10.1037/0022-3514.36.6.635

Laurin, K., Kay, A. C., and Fitzsimons, G. J. (2012). Reactance vs rationalization: divergent responses to policies that constrain freedom. *Psychol. Sci.* 23, 205–209. doi: 10.1177/0956797611429468

Legault, L., Green-Demers, I., Grant, P., and Chung, J. (2007). On the self-regulation of implicit and explicit prejudice: a self-determination theory perspective. *Pers. Soc. Psychol. Bull.* 33, 732–749. doi: 10.1177/0146167206298564

Legault, L., Gutsell, J. N., and Inzlicht, M. (2011). Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. *Psychol. Sci.* 22, 1472–1477. doi: 10.1177/0956797611427918

Macnamara, F. (2022). *University of Bradford's Gender Neutral Language Shift in Midwifery*. Available online at: https://www.thetelegraphandargus.co.uk/news/23038736.university-bradfords-gender-neutral-language-shift-midwifery/

Marks, D. (2014). What's in a name: Indian, Native, Aboriginal, or Indigenous? Available online at: https://www.cbc.ca/news/canada/manitoba/what-s-in-a-name-indian-native-aboriginal-or-indigenous-1.2784518

Marullo, S., and Edwards, B. (2000). From charity to justice: the potential of university-community collaboration for social change. *Am. Behav. Sci.* 43, 895–912. doi: 10.1177/00027640021955540

Morey, R., and Rouder, J. (2022). *BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-4.4*. Available online at: https://CRAN.R-project.org/package=BayesFactor

Munger, K. (2017). Tweetment effects on the tweeted: experimentally reducing racist harassment. *Polit. Behav.* 39, 629–649. doi: 10.1007/s11109-016-9373-5

National Assembly of State Arts Agencies (2020). *Inclusive Language Guide*. Available online at: https://nasaa-arts.org/nasaa_research/inclusive-language-guide/

O'Neil, R. M. (1997). *Free speech in the College Community*. Bloomington, IN: Indiana University Press.

Page-Gould, E., Mendoza-Denton, R., and Tropp, L. R. (2008). With a little help from my cross-group friend: reducing anxiety in intergroup contexts through cross-group friendship. *J. Pers. Soc. Psychol.* 95, 1080–1094. doi: 10.1037/0022-3514.95.5.1080

Pavey, L., Churchill, S., and Sparks, P. (2022). Proscriptive injunctions can elicit greater reactance and lower legitimacy perceptions than prescriptive injunctions. *Pers. Soc. Psychol. Bull.* 48, 676–689. doi: 10.1177/01461672211021310

Peters, G. J. Y., Ruiter, R. A., and Kok, G. (2014). Threatening communication: a qualitative study of fear appeal effectiveness beliefs among intervention developers, policymakers, politicians, scientists, and advertising professionals. *Int. J. Psychol.* 49, 71–79. doi: 10.1002/ijop.12000

Pettigrew, T. F., and Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* 90, 751. doi: 10.1037/0022-3514.90.5.751

Pinker, S. (1994). "The game of the name," in *New York Times*. Available online at: https://www.nytimes.com/1994/04/05/opinion/the-game-of-the-name.html (accessed Nov 11, 2022).

Plant, E. A., and Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: acceptance or backlash? *J. Exp. Soc. Psychol.* 37, 486–501. doi: 10.1006/jesp.2001.1478

Price, O. (2023). *University Sparks Language Row as it Advises Students to Refer to Each Other as 'They' Until the Person Reveals Their Preferred Pronouns to Create 'Culture of Inclusion'*. Available online at: https://www.dailymail.co.uk/news/article-11784439/Kent-University-sparks-woke-language-row-advice-refer-people-pronouns-unknown.html

Proulx, T., Costin, V., Magazin, E., Zarzeczna, N., and Haddock, G. (2022). The progressive values scale: assessing the ideological schism on the left. *Pers. Soc. Psychol. Bull.* 01461672221097529. doi: 10.1037/t89723-000

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Read, B. (2018). Truth, masculinity and the anti-elitist backlash against the university in the age of Trump. *Teach. Higher Educ.* 23, 593–605. doi: 10.1080/13562517.2018.1457636

Richard, F. D., Bond Jr, C. F., and Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Rev. General Psychol.*7, 331–363. doi: 10.1037/1089-2680.7.4.331

Richeson, J. A., and Shelton, J. N. (2007). Negotiating interracial interactions: costs, consequences, and possibilities. *Curr. Dir. Psychol. Sci.* 16, 316–320. doi: 10.1111/j.1467-8721.2007.00528.x

Rios, K., and Wynn, A. N. (2016). Engaging with diversity: framing multiculturalism as a learning opportunity reduces prejudice among high White American identifiers. *Eur. J. Soc. Psychol.* 46, 854–865. doi: 10.1002/ejsp.2196

Roberts, R. (2017). *Hull University Threatens to Mark Down Students Who Don't Use Gender Neutral Pronouns*. Available online at: https://www.independent.co.uk/news/uk/home-news/hull-university-gender-neutral-pronouns-students-mark-down-not-use-a7664581.html

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Math. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Rozin, P., and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psychol. Rev.* 5, 296–320. doi: 10.1207/S15327957PSPR0504_2

Saad, L. F. (2020). *Me and White Supremacy: How to Recognise Your Privilege, Combat Racism and Change the World*. Naperville, IL: Source Books.

Savani, K., Markus, H. R., and Conner, A. L. (2008). Let your preference be your guide? Preferences and choices are more tightly linked for North Americans than for Indians. *J. Pers. Soc. Psychol.* 95, 861–876. doi: 10.1037/a0011618

Shelton, J. N., Richeson, J. A., Salvatore, J., and Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychol. Sci.* 16, 397–402. doi: 10.1111/j.0956-7976.2005.01547.x

Strauts, E., and Blanton, H. (2015). That's not funny: Instrument validation of the concern for political correctness scale. *Pers. Individ. Dif.* 80, 32–40. doi: 10.1016/j.paid.2015.02.012

Tom, A., and Granié, M. A. (2011). Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. *Accid. Analy. Prevent.* 43, 1794–1801. doi: 10.1016/j.aap.2011.04.012

Vertovec, S., and Wessendorf, S. (2010). *The Multiculturalism Backlash: European Discourses, Policies, and Practices*. London: Routledge. doi: 10.4324/9780203867549

Wagenmakers, E. J., Wetzels, R., Borsboom, D., and Van Der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem. *J. Pers. Soc. Psychol.* (2011) 100, 426–32. doi: 10.1037/a0022790

Webb, T. L., and Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychol. Bullet.* 132, 249. doi: 10.1037/0033-2909.132.2.249

Wicklund, R. A., Slattum, V., and Solomon, E. (1970). Effects of implied pressure toward commitment on ratings of choice alternatives. *J. Exp. Soc. Psychol.* 6, 449–457. doi: 10.1016/0022-1031(70)90055-7

Wilson, J. K. (2020). *The Myth of Political Correctness: The Conservative Attack on Higher Education*. Durham: Duke University Press.

Wilson, T. (2011). *Redirect: The Surprising New Science of Psychological Change*. London: Penguin.

Worchel, S., Arnold, S., and Baker, M. (1975). The effects of censorship on attitude change: The influence of censor and communication characteristics. *J. Appl. Soc. Psychol.* 5, 227–239. doi: 10.1111/j.1559-1816.1975.tb00678.x