Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Employee "Free" Speech vs. Organizational Censorship on Social Media: Balancing the Tension Between Free Expression and Safeguards to Foster Psychological Safety

Brittany C. Solomon, Alexandra K. Scott, Matthew E. K. Hall

To cite this article:
Brittany C. Solomon, Alexandra K. Scott, Matthew E. K. Hall (2025) Employee "Free" Speech vs. Organizational Censorship on Social Media: Balancing the Tension Between Free Expression and Safeguards to Foster Psychological Safety. Organization Science

Published online in Articles in Advance 14 May 2025

. https://doi.org/10.1287/orsc.2024.19506

https://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Articles in Advance, pp. 1–22
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Employee "Free" Speech vs. Organizational Censorship on Social Media: Balancing the Tension Between Free Expression and Safeguards to Foster Psychological Safety

Brittany C. Solomon,[a,*] Alexandra K. Scott,[a] Matthew E. K. Hall[b]

[a] Department of Management & Organization, Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana 46556; [b] Department of Political Science, University of Notre Dame, Notre Dame, Indiana 46556

*Corresponding author

Contact: bsolomon@nd.edu, https://orcid.org/0000-0002-0462-7535 (BCS); ascott28@nd.edu, https://orcid.org/0009-0008-5225-1625 (AKS); matt.hall@nd.edu, https://orcid.org/0000-0002-0802-0415 (MEKH)

Received: September 7, 2024
Revised: November 27, 2024; March 2, 2025; March 6, 2025
Accepted: March 7, 2025
Published Online in Articles in Advance: May 14, 2025

https://doi.org/10.1287/orsc.2024.19506

Abstract. Free speech is a core principle in most democratic societies. However, the proliferation of prejudiced speech on social media has prompted intense debates about censorship. As a corollary in the workplace, and against the backdrop of online boundary blurring, we recruited a U.S. nationally representative sample to document that coworkers' prejudiced social media posts concern Democrats whereas employer censorship of such posts concerns Republicans. To investigate this tension, we integrated boundary research into the psychological safety literature to theorize that organizational censorship on social media differentially affects liberals' and conservatives' individual-level psychological safety at work. We also theorized that psychological safety may be fostered—for all employees—by exploiting the tension between safeguards and free expression in organizational social media policies. Taking a multimethod approach, we first explored reactions to organizational censorship of prejudiced speech using open-ended responses (Study 1). Next, we demonstrated in preregistered observational and experimental surveys that such censorship increases Democrats' psychological safety but decreases Republicans' psychological safety (Studies 2–3). Finally, in a preregistered, conjoint survey experiment (Study 4), we examined moderators that could help manage this tension. We found that alignment between organizational values and censorship practices, as well as censorship of anti-Black speech, threatening speech, and references to one's organization, can improve psychological safety for Democrats *without* undermining psychological safety for Republicans. As political polarization permeates societies worldwide, maintaining the social-relational fabric of organizations will require policies that balance ideological demands. We conclude by discussing implications for psychological safety, "free" speech, and organizational censorship.

History: This paper has been accepted for the *Organization Science* Special Issue on Psychologically-(Un)Safe Climates in the Age of Digital and Social Tensions.

Funding: This project was funded by the University of Notre Dame's Mendoza College of Business and Rooney Center for the Study of American Democracy.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2024.19506.

Keywords: censorship • free speech • social media • boundary blurring • boundary control • psychological safety

The current era of global democratic upheaval, combined with the ubiquity of social media, has reignited long-standing debates about free speech in society (Mitchell and Walker 2021, Boone 2023, St. Aubin and Liedke 2023). Hate speech has been a central focus of these debates (Solomon et al. 2024) because social norms that discourage prejudice can lie in tension with free speech values. Most research on free speech attitudes has focused on the general public (e.g., Kozyreva et al. 2023, Pradel et al. 2024, Solomon et al. 2024). In contrast, we emphasize that individuals often operate within organizational contexts and thus the free speech of fellow organizational members becomes especially relevant. Compared with the proliferation of prejudiced attitudes on social media platforms (Costello et al. 2016, Ronson 2016, Walther 2022), prejudice is less commonly expressed in work environments because of normative constraints (Goncalo et al. 2015, Li et al.

2024). But contexts commonly collide on social media, with coworkers and organizations both blurring the personal/professional boundary (Ollier-Malaterre et al. 2013). Given the importance of managing a diverse workforce and implementing policies that combat prejudice (Paluck et al. 2021, Torrez et al. 2024), organizations may try to limit their employees' "free" expressions of prejudice on social media to facilitate psychological safety—but should they? Since the outbreak of violence in Gaza in 2023, prominent companies like Citigroup and Delta have fired employees for antisemitic and anti-Palestinian posts on social media (Gibson 2023, Elsamra 2024). Yet little is known about the effects of such policies. We argue that, paradoxically, *both* employee free speech on social media *and* its censorship can undermine psychological safety at work. We also argue that carefully balancing free expression and organizational safeguards should help manage this tension.

Rooted in mutual trust and respect, psychological safety allows employees to feel comfortable being themselves and safe taking social risks, such as expressing personal opinions or controversial views (Kahn 1990, Edmondson 1999). Dominant themes in this literature include "getting things done, learning behaviors, improving the work experience, and leadership," especially at the group level where this construct originated (Edmondson and Bransby 2023, p. 55), but also extending to the individual and organizational levels. The literature is replete with studies showing that psychological safety improves performance and teamwork, learning and knowledge sharing, voice, creativity, job engagement, job satisfaction, and stress (for reviews see Edmondson and Lei 2014, Frazier et al. 2017, Newman et al. 2017, Edmondson and Bransby 2023). And the role of leaders is emphasized in cultivating psychologically safe work environments, for example, through feedback sharing (Coutifaris and Grant 2021), inclusive leadership, leader-member exchange, leader openness, and trust in leadership. Supportive organizational practices are also important antecedents of psychological safety (Singh et al. 2013, Newman et al. 2017).

One organizational practice that may aim to improve psychological safety is censoring employees' prejudiced social media posts. But free speech has always been controversial (Boone 2023). As such, we suspect censorship can also undermine psychological safety if employees feel constrained in their freedom to be themselves and express their views (Brehm and Brehm 2013), especially in a nonwork context like on public social media. We argue that organizations face a classic moral dilemma in crafting social media policies that pits protecting employees from prejudiced speech (i.e., control vis-à-vis organizational censorship) against preserving employees' free expression (i.e., autonomy vis-à-vis protecting "free" speech). Thus, organizations must adopt policies that manage these competing interests.

Unfortunately, crafting social media policies is complicated because different employees likely react to censorship in very different ways. Among the general public, liberals are now consistently more censorial than conservatives, who tend to champion free speech (Chong et al. 2024). Prior to developing our formal expectations regarding psychological safety, using a U.S. nationally representative probability-based sample (the gold standard for external validity in a population), we explored and found support for two propositions related to online boundary blurring and control, respectively: A significant portion of the population has (a) seen a coworker post something on social media that made them concerned their coworker was prejudiced and/or (b) been concerned their employer might censor them for posting something supposedly prejudiced on social media. We also uncovered a compelling tension: Democratic employees are more concerned about seeing prejudiced coworker posts, whereas Republican employees are more concerned about censorship by their employers (see the Online Appendix).

Drawing on ideologically based political differences (Jost et al. 2009, Swigart et al. 2020, Jost 2021), we position the current investigation at the understudied intersection of research on individual-level psychological safety and online boundary management (Rothbard and Ollier-Malaterre 2016, Behrend et al. 2024). We theorize that organizational censorship policies, a form of boundary control, differentially influence psychological safety because of ideological disagreement (Stanko and Beckman 2015). As we will explain, such policies instill norms regarding what sentiments are versus are not acceptable to express. Organizations that censor employees for prejudiced social media posts should signal to liberals—who prioritize safeguards—that their concerns and opinions related to inclusivity and social equity are aligned with the organization's values and can be expressed freely without fear of being dismissed or disparaged (e.g., labeled "woke"). But conservatives—who prioritize free speech—should feel less psychologically safe, as policies that exert such control in nonwork contexts marginalize their values.

But how can such tension be managed to foster psychological safety in pluralistic organizations? We posit that maximizing psychological safety must involve more than merely deciding whether to censor employees. And we emphasize the value of control *and* autonomy, and thus in rejecting "the conventional and divisive either/or approach to managing competing forces" (Graetz and Smith 2008, p. 276). By exploiting both sides of this tension, organizations can craft social media policies to increase psychological safety for liberals *without* decreasing psychological safety for conservatives. We propose four key factors (organizational values, the targeted group, speech severity, and company affiliation) that can each moderate the

effects of censorship on psychological safety to create this equilibrium.

Taken together, our first goal is to document a modern organizational tension: Censorship of social media content differentially affects the psychological safety of liberal versus conservative employees. Our second goal is to develop novel theorizing, which can be leveraged in practice, for crafting an ideological consensus to balance safeguards with free expression. Embracing this tension should foster an adequate psychological safety equilibrium for employees on both sides. We employ a multimethod investigation, including the collection of exploratory and descriptive data (prestudy and Study 1) and a series of preregistered, observational and experimental surveys (Studies 2, 3, and 4).

Now a mainstream topic, psychological safety has "come of age," often operating as a mechanism (Edmondson and Bransby 2023, p. 69) or moderated predictor of various outcomes (Newman et al. 2017). But, according to Edmondson and Bransby (2023, p. 71), how to create psychological safety remains "the most glaring gap in the literature." We highlight that efforts to create psychological safety can also subvert this goal. Thus, we add to research on the boundary conditions of antecedents to psychological safety (e.g., Harvey et al. 2019, Tu et al. 2019, Rego et al. 2021), in particular antecedents to individual-level psychological safety (e.g., Liu et al. 2017, Jiang et al. 2019, Dwivedi et al. 2023). Prior work shows that different work groups within an organization often differ in their sense of psychological safety, for example, because of *differences* in their leadership (Edmondson 1996, 2003). In our research, we highlight the tension created when the *same* censorship policy fosters psychological safety for some individuals (i.e., liberals) while undermining psychological safety for others (i.e., conservatives). As such, we focus on differences between social groups, which prior work in this arena has largely overlooked—even regarding commonly studied demographics, such as race and gender (cf. Singh et al. 2013, Dwivedi et al. 2023). Our work also answers calls to further investigate organizational- and individual-level interactions (Edmondson and Lei 2014, Newman et al. 2017) and triangulate data, using a multimethod approach, to provide a deeper understanding of both experiences and causal relationships (Edmondson and Lei 2014).

Although psychological safety has been tethered to most mainstream organizational behavior topics, it has remained largely separated from research on boundary management (Edmondson and Bransby 2023) and societal politics (cf. Leigh and Melwani 2022). Here, we also contribute to the psychological safety literature by integrating research on online boundary management (Rothbard and Ollier-Malaterre 2016, Ollier-Malaterre et al. 2019, Behrend et al. 2024) and the growing body of organizational research on ideology/partisan affiliation (see, e.g., Swigart et al. 2020). Prior research on online boundary blurring has focused on coworker connections and consequences (Ollier-Malaterre et al. 2013, Pillemer and Rothbard 2018, Rothbard et al. 2022). Here, we reason that online boundary blurring has implications for psychological safety, and we examine how organizational boundary *control* affects the psychological safety of liberal versus conservative employees. We also focus on prejudiced social media content, which is largely unexplored in the online boundary blurring literature. These contributions are important given the prominence of untailored self-disclosures to online audiences that lack workplace norms (Pillemer and Rothbard 2018) and debates about free speech versus censorship (Chong et al. 2024). Also, building on studies of organizational social media surveillance and policies (Stohl et al. 2017, Banghart et al. 2018), we argue that censoring employee speech on personal social media accounts influences psychological safety *at work*. Thus, in addition to established antecedents that are traditionally associated with the workplace, such as leadership behaviors, coworker support, and work design characteristics (Frazier et al. 2017, Newman et al. 2017, Edmondson and Bransby 2023), we emphasize that individual-level psychological safety can be shaped across the work/nonwork boundary. Finally, most microlevel research on political affiliation highlights the role of basic in-group/out-group dynamics, such as in employment decisions (Roth et al. 2020, Mönke et al. 2024). Rather than focusing on psychological safety as a reaction to in-party versus out-party coworkers, we add to that literature by examining partisans' asymmetric sense of psychological safety in response to the same organizational policy. Additionally, we join the broader conversation on manifestations of ideology and societal politics in the workplace (e.g., Patil and Bernstein 2022, Rheinhardt et al. 2024, Wowak and Busenbark 2024).

Our work also has clear practical implications for fostering psychological safety in a politically polarized society. For example, our theory and findings point to the importance of questioning policies that are assumed to generally promote psychological safety across employees and identifying, rather than ignoring, ideologically infused tensions that pit employees' preferences against each other. They also point to the value of creating an equilibrium between safeguards and free expression to avoid backlash and effectively manage ideological demands. We expound upon these implications in the Discussion.

Exposure to Prejudice and Censorship via Online Boundary Blurring and Control

Whether one integrates or segments a given personal/professional boundary is fundamental to boundary theory (Nippert-Eng 1996, Ashforth et al. 2000, Rothbard et al. 2005). And any connection to one's colleagues

online—intentional or inadvertent—indicates some degree of integration (Rothbard and Ollier-Malaterre 2016). Although people take different approaches to configuring privacy settings, with whom they connect, and how they present themselves (Ollier-Malaterre et al. 2013, Schlosser 2020, Rothbard et al. 2022), effectively managing personal/professional boundaries online requires considerable awareness, motivation, and skill (Ollier-Malaterre et al. 2019). Thus, despite calls to reconsider the value of segmentation (Creary and Locke 2022), especially in today's digital society (Rothbard and Ollier-Malaterre 2016), online boundary blurring has become normalized. And it is almost inevitable for many workers around the world (Ollier-Malaterre et al. 2019), contributing to the culture of social surveillance in which people keep tabs on one another (Lyon 2018) at work (Patil and Bernstein 2022) or not (Olmsted et al. 2016). Despite impression management efforts, any behavior on social media inherently reveals information about oneself (Pillemer and Rothbard 2018, Schlosser 2020). Self-disclosure is generally linked to positive social relations (Collins and Miller 1994). But one cannot tailor their image on social media to certain spectators as they would in most contexts (Behrend et al. 2024). Thus, it is not surprising that navigating boundaries on social media may positively *or* negatively affect coworker friendships (Pillemer and Rothbard 2018) as well as liking and respect in professional relationships, depending on one's approach (Ollier-Malaterre et al. 2013)—and what aspect of their identity they express.

Boundary blurring on social media may lead to a plethora of uncurated disclosures that coworkers might never encounter otherwise (Pillemer and Rothbard 2018, Behrend et al. 2024). Social media has been criticized for outing those with sensitive identities (Cho 2018, Lauriano and Coacci 2023), but we highlight that it can also "out" coworkers who harbor prejudice. People tend to express themselves more freely on social media than at work (Costello et al. 2016, Ronson 2016), where bias suppression norms and the consequences for violating them are more psychologically proximal (Crandall and Stangor 2005, Goncalo et al. 2015, Li et al. 2024). Of course, some people adopt coded language on social media (Al-Kire et al. 2024), and overt and covert slurs and prejudice still exist in the workplace (Rosette et al. 2013, Hebl et al. 2020). Nonetheless, we contend that employees are especially likely to learn of coworker prejudice through boundary blurring on social media. Like detecting warmth and competence online (Rothbard et al. 2022, Roccapriore and Pollock 2023), we suspect employees are quite attuned to signals of prejudice, whether it is because they endorse antiprejudiced values or harbor prejudice themselves.

We argue that organizations must contend with potential spillover from prejudiced employee speech on social media to the workplace. In contrast to boundary work enacted by individuals (e.g., Ramarajan and Reid 2013), organizational boundary control pertains to the organization's management of the personal/professional boundary (Thomas and Ganster 1995, Kossek and Lautsch 2012, Stanko and Beckman 2015). Boundary control is often aimed at influencing when, where, and how employees spend their time in light of their work and nonwork roles (Perlow 1998). Like the broader function of organizational control, boundary control is more directly aimed at affecting employees' work in their job role (Ouchi and Maguire 1975, Ocasio and Wohlgezogen 2010). Here, we focus on how boundary control influences psychological safety, which should in turn influence one's work (Edmondson and Bransby 2023). On social media, employees are apt to use speech that reflects their personal identity as members of society rather than their professional identity as members of their organization (Stohl et al. 2017, Banghart et al. 2018, Neeley and Leonardi 2018). And it is through organizations' social media policies that organizations are most likely to breach the online personal/professional boundary. Policies that censor prejudiced speech on social media inherently exert control over employees' non-work-related expression. In their absence, organizations allow employees to maintain autonomy over their nonwork speech.

Based on an "invasive boundary logic," many of the world's largest corporations recognize that separating personal/professional roles on public social media platforms is untenable (Banghart et al. 2018; see also Hidy 2020). Business leaders also tend to consider social media surveillance a best practice (Forbes Human Resources Council 2017). And organizations increasingly implement social media policies that tend to hinder rather than promote employee "free" speech (Stohl et al. 2017). Employees commonly but mistakenly believe that their employers cannot legally punish them for discriminatory and disparaging posts on their personal social media accounts (Graves 2021). But, in reality, such online speech and "personal" online activities fall within an employer's authority if they violate company policy. In practice, censorship arises through either active monitoring (Karunakaran et al. 2022) or reports by coworkers and other stakeholders (Hidy 2020). To illustrate, Dell Technologies "has zero tolerance for racism, bigotry, misogyny, express or implied threats of harassment or physical harm, or hate speech," and its policy alerts employees to "assume that anything you do on social media—whether on a business or personal account—could be viewed by a colleague, supervisor, partner, supplier, competitor, investor, customer or potential customer" (Dell Technologies 2023). Organizations largely censor employees for preemptive risk management purposes, such as safeguarding their reputation, operations, and

proprietary information (Hidy 2020). Here, we suggest that their social media policies regarding prejudiced speech may also affect employees' psychological safety—but not in a uniform way for all employees.

Censorship and Differential Psychological Safety

Recall that psychological safety at work reflects one's sense of vulnerability concerning others' reactions to the expression of one's thoughts and concerns, and comfort in taking such social risks (Edmondson 1999, 2018). Psychological safety is less likely to vary within than across groups in an organization, largely because of group-level differences, for example, in leadership behaviors (Edmondson 1996, 2003). But how does the same organizational practice, like a social media censorship policy, affect psychological safety when employees systematically differ in their values and in a society where values-based differences are prominent? Supportive organizational practices help cultivate psychological safety at the individual level (Newman et al. 2017). But it would be imprudent to assume that ostensibly supportive practices, like censorship, have consistently positive or negative effects across employees with different political ideologies. As politics now polarizes most advanced, democratic societies more so than historically divisive characteristics, such as race and religion (Silver et al. 2021), we focus on political ideology as a critical yet understudied social group difference that may influence psychological safety.

We theorize that censorship of employees who post prejudiced content on social media fosters the psychological safety of liberals while impeding the psychological safety of conservatives. Prior work has examined individual differences as predictors of boundary management preferences and behaviors (Rothbard and Ollier-Malaterre 2016). Here, we build on such work by positing that political ideology conditions reactions to boundary *control* in the form of social media censorship. Given the inevitability of tensions in pluralistic organizations (Kraatz and Block 2008, Smith and Lewis 2011), we draw on well-established ideological differences (Jost et al. 2009, Jost 2021) to develop this expectation.

Liberals, who were traditionally more supportive of free expression, have grown wary of damage that certain rhetoric can inflict, including hate speech that further marginalizes historically disadvantaged groups (Chong et al. 2024). And their inclination toward valuing equality over hierarchy and social change over tradition (Jost et al. 2009, Swigart et al. 2020, Jost 2021) should align with prioritizing safeguards over free speech in the face of coworker prejudice. Anyone in the workplace can potentially affect others' psychological safety (Edmondson and Bransby 2023). In light of the personal/professional online boundary, we suspect that liberals are less willing or able to compartmentalize coworker prejudice on social media, potentially disrupting their psychological safety at work. But given liberals' tendency to emphasize the situational context over agency (Jost et al. 2009, Jost 2021), they should feel psychologically safer in organizations that manage online prejudice. Just as organizational punishment is enacted to instill norms (Villatoro et al. 2014), such censorship policies would signal to liberals that they can voice concerns, share opinions, ask for help, or simply be themselves at work without being dismissed as "woke," a "snowflake," or otherwise out of step with the company's prevailing values.

We reason that, if liberal employees believe that their organization disregards prejudiced speech, they will feel psychologically unsafe bringing up diversity-related or socially sensitive ideas or questions for fear of risking exclusion, strained relationships, or reputational damage like being labeled a "troublemaker." These expressions could be specific to the work itself, such as voicing unpopular concerns about an offensive marketing strategy or scrutinizing the lack of input from underrepresented stakeholders in product development. Or they could pertain to the broader organizational environment, such as asking difficult questions about progress toward diversity goals, bias in performance reviews, and data regarding pay disparities. Failing to share controversial ideas and ask uncomfortable questions to avoid these social risks could undermine learning, performance, or—of particular relevance to our research—the work experience (e.g., by creating a noninclusive environment and discouraging employees from being themselves (Edmondson and Bransby 2023)). Even if one's own social group is not targeted, they may still feel psychologically unsafe if their organization is indifferent to posts endorsing Nazis or the Ku Klux Klan, as extreme examples. But organizational censorship of antisemitic and anti-Black speech, whether through post removal or employee suspension/termination, should make liberals feel less vulnerable and more comfortable raising issues or offering suggestions that align with their values.

Additionally, liberals may hold an ideologically infused psychological contract (Thompson and Bunderson 2003) in which their organizations take responsibility for addressing employee prejudice, regardless of where it is expressed. Third parties who interpret workplace mistreatment as a moral violation expect the organization to enact punishments if they are not in a position to do so (O'Reilly and Aquino 2011). In boundaryless organizations, employees can socially sanction their coworkers on social media (Frey et al. 2023). But we contend that doing so is unlikely to improve their psychological safety at work and may provoke retaliation. Moreover, given that punishment

tends to change employee behavior (Hersel et al. 2019), censorship on social media could help prevent prejudice from spreading in addition to improving psychological safety for liberals. Thus, liberals may view censorship as a constructive and necessary form of organizational control to address prejudice and manage psychological safety.

In contrast, conservatives, who previously sympathized with censorship in order to preserve traditional social mores, now tend to embrace free speech as a justification for expressing what their critics condemn as politically incorrect ideas (Chong et al. 2024). Thus, they often decry censorship as "woke" policies that silence them and coddle "snowflakes" sheltered from disfavored opinions (Johnson 2024). We contend that conservatives' inclination to value tradition over social change and agency over situational factors (Jost et al. 2009, Swigart et al. 2020, Jost 2021) should align with prioritizing free speech over safeguards. Presumably, as defenders of free speech, conservative employees tend to embrace segmentation between personal and professional life, disliking that their personal social media activity falls within the company's purview (see, e.g., Rothbard et al. 2005, Ramarajan and Reid 2013). As such, they may be more prone to "collapse denial," in which they ignore the blurriness of online boundaries, ostensibly allowing them to behave differently online and at work (Lauriano and Coacci 2023). Although some coworkers will see this disparity, conservative employees should, nonetheless, feel psychologically safer at work when the organization does not exert control in nonwork contexts like social media.

Conservatives may also be less concerned with censorship of prejudiced speech because of ideological differences in assumptions of vulnerability (Womick et al. 2024). Liberals tend to assume that those outside the dominant social group are especially vulnerable to harm. Conservatives tend to deny this victimization of the "othered" (Gray and Kubin 2024), and their alignment with social dominance orientation and social hierarchy (Pratto et al. 1994, Jost et al. 2009) may also underlie their relative tolerance, on average, of minority-focused hate speech (Bilewicz et al. 2017). Although prior work has shown that free speech construed as a value can be used to justify racism (White and Crandall 2017), we emphasize that conservatives' support for free speech is not necessarily rooted in racist attitudes (see Solomon et al. 2024). Nonetheless, their resistance to political correctness (Rosenblum et al. 2020) might render them more likely to overlook or fail to understand how their online speech is received by others in their ostensibly nonwork lives. For example, a Victoria's Secret Stores manager was fired for a Facebook post "depicting a person wearing a Ku Klux Klan-reminiscent white, hooded robe … and was captioned 'Game 5 in LA is Free Sheet Night …'" and a picture "of an African-American female named 'Airwrecka McBride' … with the caption 'I've been spelling Erica wrong my whole life.'" She argued her posts were misconstrued and were "not actually offensive" (Hidy 2020). If conservatives worry more about mistakes and misunderstandings on social media being censored by their employer, it stands to reason that they would also worry more about making mistakes, voicing concerns, sharing opinions, or simply being themselves at work. Some policies even dismiss the possibility of innocence or misinterpretation, which could further instill fear. Again, using Dell's policy as an illustration, it states that prejudice "may be hidden in messaging … You are expected to know the meaning and importance of what you engage with on social media and how it could be received by others" (Dell Technologies 2023). Because corporate social media policies are often subjective and ambiguous, one review argues that organizations should reject civility codes in these policies in favor of employees' "right to be forgiven" for offensive social media posts (Hidy 2020). Of course, censorship does not easily comport with this "right."

Therefore, we reason that conservative employees may hold an ideologically infused psychological contract (Thompson and Bunderson 2003) in which their organizations respect their privacy and right to free expression in nonwork contexts. Even censorship that is interpreted as socially beneficial can be perceived as a significant threat to one's freedom and self-perception as an autonomous actor (Graupmann et al. 2012, Rosenberg and Siegel 2018). Thus, consistent with psychological reactance theory, which posits that restricting freedom is aversive (Brehm 1966, Brehm and Brehm 2013), threats to employee free speech on social media should undermine psychological safety for conservatives.

In light of the ideological tension we have explicated, we hypothesize the following interaction:

Hypothesis 1. *Organizational censorship* (a) *increases psychological safety for liberals and* (b) *decreases psychological safety for conservatives.*

Strengthening Effects for Liberals and Weakening Effects for Conservatives

Given the increasingly divided sociopolitical climate and recent episodes of political unrest worldwide, we highlight that organizations must not only recognize but also grapple with competing stakeholder preferences during periods of instability (Quinn and Cameron 1988). We suggest that maximizing psychological safety requires identifying areas of cross-party consensus but not in the form of agreement, per se. Rather, we theorize that certain aspects of censorship will moderate its effects on psychological safety by strengthening

the positive effect for liberals while weakening the negative effect for conservatives. Below, we delineate four factors that we expect to help strike this equilibrium: organizational values, the targeted group, speech severity, and company affiliation. The symbiotic balance we propose relies on recognizing the merits of safeguards *and* free expression (control *and* autonomy).

Organizational Values and Value Congruence

First, an organization's values are likely to play an important moderating role because they shape employees' expectations regarding its policies (Posner et al. 1985). The manifestation of values in censorship policies is a stable form of boundary control (Ocasio and Wohlgezogen 2010). Because of value congruence (Posner 2010), liberal employees likely feel psychologically safer when the organization explicitly commits to prioritizing safeguards. And conservatives likely feel psychologically safer when the organization explicitly commits to prioritizing the free exchange of ideas. However, our interest lies in identifying how organizations can craft censorship policies to maximize psychological safety for all employees. Inconsistencies between what people do and what they say they value are common in organizations (Simons 2002, Babu et al. 2020, Wagner et al. 2020), prompt perceptions of hypocrisy (Barden et al. 2005, Graham et al. 2015), and can backfire (Cha and Edmondson 2006, Greenbaum et al. 2015), including diminished psychological safety (Leroy et al. 2012, Halbesleben et al. 2013). Given the core psychological motivation for congruence between beliefs and behaviors (Festinger 1962, Elliot and Devine 1994), organizations should avoid appearing hypocritical (Carlos and Lewis 2018). Thus, censorship should especially undermine conservatives' psychological safety when organizations endorse employee free speech and, nonetheless, implement censorship. Here, we expect for censorship's negative effect on conservative employees to be weaker and nonsignificant when an organization's values are not contradicted by its actions.

Ultimately, when organizations commit to antiprejudice rather than pro-free-speech values and implement censorship accordingly, liberals should feel especially psychologically safe because of coherence between their own and the organization's values. And conservatives should feel less psychologically unsafe because of coherence in the organization's values and practices, in contrast to the hypocrisy that would emerge had the organization espoused pro-free-speech values yet enacted censorship. Thus, we expect:

Hypothesis 2a. *Organizational censorship by a company with antiprejudice (versus pro-free-speech) values moderates the effects of organizational censorship on psychological safety, such that the positive effect is stronger for liberals and the negative effect is weaker for conservatives.*

The Targeted Group and the "Overton Window" of Organizational Censorship

Next, we theorize that the targeted group in prejudiced posts will moderate the effects of censorship, again increasing psychological safety for liberals without impinging on psychological safety for conservatives. People tend to denounce behaviors that they judge as harmful to others (Schein and Gray 2018) and as targeting victimized social groups (Gray and Kubin 2024). And, by extension, they should feel psychologically safer when such behaviors are censored. Despite ideological disagreements about which groups are more or less vulnerable to victimization (Womick et al. 2024), Americans (like people in many countries) are generally less tolerant of hate speech that targets groups based on race, gender, and religion than in the past (Chong et al. 2024). And, despite misperceptions, they often agree on which targets (e.g., Blacks versus Whites) warrant more or less protection in terms of censorship on social media (Solomon et al. 2024). Here, we theorize that censorship of posts targeting marginalized groups that have historically received more (versus less) attention and protections in society should be more likely to strike an adequate balance of psychological safety between liberals and conservatives. Specifically, we focus on the comparison between Blacks and transgender people given their significant roles in social discourse.

Although liberals perceive substantially more discrimination against these two groups than conservatives do (Daniller 2021, Parker et al. 2022), we suspect that there is an "Overton window" for organizational censorship based on which groups are socially (un)acceptable to target. Conceptually, the Overton window represents the range of ideas and policies that are considered acceptable and mainstream; those that fall outside the window are considered too radical and less likely to gain traction or be implemented (Astor 2019). Expressing antitransgender ideas is *relatively* more common and—for *some* people—more socially acceptable, whereas expressing anti-Black ideas is *relatively* less common and—for *more* people—less socially acceptable. Thus, we expect censorship of anti-Black (versus antitransgender) speech to increase psychological safety for liberals *without* diminishing psychological safety for conservatives. In other words, liberal and conservative employees can agree that anti-Black speech falls outside of the Overton window, and thus its censorship should be more acceptable to those across the ideological spectrum. In contrast, employees are more likely to disagree about the social acceptability of antitransgender speech (Parker et al. 2022), and thus its censorship might tend to make liberals feel psychologically safer but reduce psychological safety for conservatives.

Although an organization's values tend to be stable, we note that the Overton window is inherently

evolving. The Overton window reflects the public's current opinion, but it shrinks, expands, or shifts with changes in popular attitudes. Importantly, whether it is socially acceptable to aim prejudice at certain social groups changes over time. For example, antigay/antilesbian attitudes that were previously common and condoned in society are now broadly condemned (Norris 2023, Brenan 2024). Similarly, although debates over transgender status and rights are now considered mainstream, the public may one day agree that anti-transgender arguments are unacceptable. As such, popular attitudes toward the acceptability of speech targeting a particular group represent the potential for change, which has implications for how psychological safety may be cultivated over time; see our Discussion.

As a point of comparison, in this age of political polarization, antiliberal and anticonservative rhetoric is explicit and socially acceptable (Mason 2018, Iyengar et al. 2019) yet pertains to groups that are not regarded as marginalized or in need of safeguarding. Therefore, despite in-party favoritism (Amira et al. 2021), we do not expect employees to feel psychologically safer when organizations censor speech targeting those with similar ideological views. Focusing again on safeguards as we did in Hypothesis 2a, we expect:

Hypothesis 2b. *Organizational censorship of anti-Black (versus antitransgender, antiliberal, or anticonservative) posts moderates the effects of organizational censorship on psychological safety, such that the positive effect is stronger for liberals and the negative effect is weaker for conservatives.*

Speech Severity, Company Affiliation, and Escape Valves

Having focused on safeguards in terms of organizational values and the targeted group, we now focus on creating a balance with free expression. In particular, we examine speech severity and company affiliation as straightforward moderators that, we contend, provide social escape valves for employee speech. These escape valves are important because they help ensure an adequate degree of autonomy given online boundary control.

First, recent work has shown that both liberals' and conservatives' support for censoring toxic and hate speech on social media depends on its severity (Pradel et al. 2024, Solomon et al. 2024). The adverse effects of coworker incivility have been well established (Cortina et al. 2017). But expecting employees to adhere to organizational codes of civility on their personal social media accounts may be intrusive and unrealistic (Hidy 2020). Thus, building on findings in prior work (Pradel et al. 2024, Solomon et al. 2024), we suspect that censoring only threatening (versus merely critical or even insulting) posts will help ensure psychological safety for both liberals and conservatives. As such, those who are prone toward prejudiced expression still have a censorship-free avenue, via less severe, nonthreatening language, to express these attitudes without violating the organization's censorship policies.

Regarding company affiliation, whether an employee's social media account references their organization should also moderate the effects of censorship. Censorship of work-related posts, which are similar to referencing one's organization, is viewed as fairer than censorship of non-work-related posts (Parker et al. 2022). It may seem obvious that one can simply refrain from affiliating their account with their organization (Ollier-Malaterre et al. 2013). However, employees often face challenges in navigating the online personal/professional boundary regardless of their job-specific technological proficiency. For instance, digital cultural capital is a new form of cultural capital that includes "awareness of the need to actively monitor one's use of technology and the motivation and skill to do" (Ollier-Malaterre et al. 2019, p. 436). Such capital is considered "a major asset for people who want to exert individual agency" in response to increasing organizational control in this digital age (Ollier-Malaterre et al. 2019, p. 439). As such, having the option to post online provided that the posts are not affiliated with one's organization should provide another social escape valve that helps preserve one's freedom of expression and, importantly, should strengthen the positive effect and weaken the negative effect of censorship for liberals and conservatives, respectively:

Hypotheses 2c and 2d. *Organizational censorship of* (c) *threatening (versus critical or insulting) posts and* (d) *accounts that reference (versus do not reference) the poster's organization each moderates the effects of organizational censorship on psychological safety, such that the positive effect is stronger for liberals and the negative effect is weaker for conservatives.*

Overview of Studies

We first conducted an open-ended survey to explore various aspects of organizational censorship that may shape responses to the practice (Study 1). Then, to test Hypothesis 1, we conducted an observational survey (Study 2) and a survey experiment (Study 3) assessing the effects of organizational censorship on partisans' psychological safety. Finally, to test Hypotheses 2a–d, we conducted a conjoint, survey experiment to examine moderators that may help organizations achieve an equilibrium that maximizes psychological safety for both liberal and conservative employees. Note that the qualitative responses in Study 1 enabled us to gather reactions to organizational censorship without superimposing our own preconceptions, and the observational survey in Study 2 prioritized external validity by asking about participants' actual personal

experiences. Moreover, Studies 3 and 4 of our multimethod approach align with the growing recognition by organizational scholars that experiments have been underutilized. Here, we answer the call to employ experimental designs to test and develop theory, facilitating causal inferences, internal validity, and reproducible research while embracing simplicity (Levine et al. 2023).

All studies were limited to full-time employees in the United States to avoid unnecessary complexity related to different partisan and ideological labels, political cultures, and societal norms across national boundaries. We recruited each sample through CloudResearch Connect, quota-matched to the U.S. census based on race and gender and evenly balanced between Democrats and Republicans. According to the political science literature, the vast majority of liberals and conservatives identify and behave as Democrats and Republicans, respectively (Levendusky 2009). And political polarization in the United States (like many countries) manifests in a binary fashion, despite the large share of nonvoters and supposed independents. Nonvoters and voters are evenly divided on the favorability of the two major parties, most major policy issues, and key political figures (Knight Foundation 2020). And independents overwhelmingly *lean toward* one of the major parties and behave like partisans (Keith et al. 1992, Hawkins and Nosek 2012, Pew Research Center 2014). Thus, we excluded pure independents and coded leaners toward a political party as partisans (Hawkins and Nosek 2012). We also excluded those who failed either of two standard attention checks. Exclusions occurred during data collection and did not count toward our sample sizes. Sample parameters, hypotheses, conditions, and measures for Studies 2–4 were preregistered, and the data, syntax, and Institutional Review Board (IRB) approvals are posted on the Open Science Framework (OSF). See the Online Appendix for the OSF link, links to preregistrations (which include power analyses), survey wording, and regression tables. Note that we preregistered an original three-item scale of psychological safety based on extant definitions of the construct (Kahn 1990, Edmondson 1999). But in the main text, we present results for Studies 2 and 3 using an adapted version of Edmondson's (1999) seven-item index. In both studies, the scales were highly correlated and results were robust to either scale (see the Online Appendix). For Studies 2–4, we controlled for age, gender, and race. All p-values are based on two-tailed tests.

Study 1

Although some studies have analyzed the content of organizational social media policies (Stohl et al. 2017, Banghart et al. 2018), it is unclear which aspects of censorship underlie employees' reactions to the practice. Thus, in Study 1, we sought to identify specific aspects of an organization's social media policy that may ultimately influence employees' psychological safety in response to that policy. To do so, we explored reasons employees support or oppose such policies, which types of statements should or should not be censored, and particular types or aspects of censorship that should or should not be allowed—all without directing or limiting the participants' responses. We recruited 103 participants to share their perceptions and attitudes about censorship in their respective organizations.

Measures

Censorship Perceptions. Participants reported whether they thought the company they currently work for would do something to censor an employee who posted a statement on social media that was prejudiced against a variety of racial, gender, LGBTQ (lesbian, gay, bisexual, transgender, queer/questioning), religious, and political groups (1 = *Yes*, 0 = *No*).

Censorship Support. Participants reported whether they thought their company should censor employees who post prejudiced statements on social media (1 = *Yes*, 0 = *No*; *N*'s = 74 and 29).

Censorship Reasons. Participants answered an open-ended question asking why they thought their company should or should not censor employees (based on their reported censorship support) and what impact they thought such presence or absence of censorship would have.

Censorship Targets. Participants answered an open-ended question asking what types of prejudiced statements on social media should or should not prompt their company to censor its employees (based on their reported censorship support). Note we did not ask them to name specific groups.

Censorship Specifics. Participants answered an open-ended question asking what specifically would be appropriate or inappropriate for their company to do to censor its employees for prejudiced statements on social media (based on their reported censorship support).

Other Factors. Participants reported what other factors their company should take into consideration when setting its social media policy.

Partisan Affiliation. We coded participants' *partisan affiliation* based on which major political party they identified with or leaned toward (1 = *Republican*, 0 = *Democrat*).

Results

We used an iterative process to identify emergent themes and create categories for coding common responses. The first and second authors independently coded responses to the open-ended questions regarding censorship reasons, targets, specifics, and other factors. The 14% of disagreements that emerged were adjudicated by the third author.

Participants' perceptions of whether their organization would censor prejudiced speech on social media varied depending on the targeted group. Participants were most likely to report that their company would censor speech targeting Blacks (74%), gays/lesbians (71%), Asians (70%), and Hispanics (70%). Participants were least likely to report that their organization would censor speech targeting Democrats (41%), Republicans (44%), and Whites (45%). Regarding censorship support, most participants (71%) indicated that their company should censor prejudiced social media posts, but support was higher among Democrats (79%) than among Republicans (61%) ($t(101) = 2.01$, $M_{\text{diff}} = 0.18$, $SE = 0.09$, $d = 0.41$, $p < 0.05$).

The most common reason reported for supporting organizational censorship of prejudiced content was to protect the company's public image or avoid negative press (38 of 74). For example, one participant explained, "The company needs to censor employees to protect its own branding and reputation." The next most common response was to foster a positive organizational culture by promoting respect, morale, inclusivity, cooperation, and trust (26 of 74). In fact, seven participants specifically discussed the importance of censorship for cultivating a "safe" climate. For example, one participant wrote, "I think the employee should [be] censor[ed] in order to create more of a safe work environment." And 12 participants indicated that the company should censor prejudiced posts to make it clear that the company does not condone the behavior (e.g., "I believe the company's censoring them sends a clear message that the idea or posting of it is incorrect and will not be tolerated by the organization").

The most common reason for opposing organizational censorship of prejudiced content was to protect free speech rights (18 of 29). For example, one participant asserted, "People should have the right to free speech," and another explained, "This would infringe on our freedom of speech." Participants also reported an interest in segmenting their private life from their work life (12 of 29). For example, one participant wrote, "I think there should be a line between work life and personal life. If the employee did this on their own time on a personal social media account, it's none of the company's business."

Supporters of organizational censorship spontaneously emphasized the importance of censoring speech that targeted marginalized groups, such as racist (38 of 74), sexist (19 of 74), and homophobic (19 of 74) speech. And many censorship supporters (37 of 74) as well as some opponents (5 of 29) indicated support for censoring generally prejudiced, discriminatory, or hateful speech. Some censorship supporters (9 of 74) and opponents (4 of 29) also endorsed censoring violent speech. Additionally, some censorship opponents would support censoring speech that referenced the company or its employees (5 of 29). However, many censorship opponents opposed censoring any speech whatsoever (12 of 29).

Next, participants varied widely in their opinions regarding appropriate types of organizational censorship. Many censorship supporters preferred that the organization simply issue a warning (19 of 74) or require the offender to remove the prejudiced post (32 of 74). Others suggested the company should reiterate its social media policy (10 of 74) or require the offender to complete a training program (13 of 74). But many supporters preferred a more severe punishment, such as suspension or termination (30 of 74). And others suggested that the severity of the punishment should vary depending on the severity or repetition of the offense (18 of 74). However, some opponents of censorship especially objected to human resources training (5 of 29) or termination (7 of 29) as a punishment for prejudiced posts.

Finally, when offered the opportunity to report other factors their organization should consider when creating its censorship policy, participants listed the importance of free speech (11 of 103), the company's values (12 of 103), whether the posts referenced the company or its employees (12 of 103), and the importance of having a clear policy (12 of 103). But most responses to this item were left blank.

In sum, the spontaneous, open-ended responses indicated that, in crafting social media policies, organizations must balance concern for promoting a positive organizational climate against protecting employees' freedom of speech and private life. In particular, an organization should consider alignment between its censorship policy and its values, the group targeted in the employee's social media post, the severity of the speech, and whether the post references the organization. Additionally, organizations should consider censoring prejudiced posts by either requiring their removal or suspending/terminating the employee. Notably, participants' responses also confirmed the safeguards/free expression tension we theorized, which would otherwise remain obscure or be subject to demand effects with exclusively quantitative data (Turner et al. 2017).

Study 2

In Study 2, we assessed whether employee expectations regarding their own organization's censorship of prejudiced social media activity are differentially

associated with psychological safety for Democratic versus Republican employees (Hypothesis 1). We recruited 704 participants. We had some difficulty recruiting Republican women of color, which is common given partisan demographic patterns. Therefore, the final sample for Study 2 leaned slightly more Democratic (59%) than Republican (41%).

Measures

Psychological Safety. We used Edmondson's (1999) seven-item psychological safety index ($\alpha = 0.91$) on a seven-point Likert-type scale from one (*strongly disagree*) to seven (*strongly agree*). We adapted the scale to refer to the individual participant's feelings and replaced one item; see the Online Appendix for details.

Organizational Censorship. We coded participants' reports of whether their organization would do something to censor employees who post a prejudiced statement on social media (1 = *Yes*, 0 = *No*).

Partisan Affiliation. We coded *partisan affiliation* in the same manner as in Study 1.

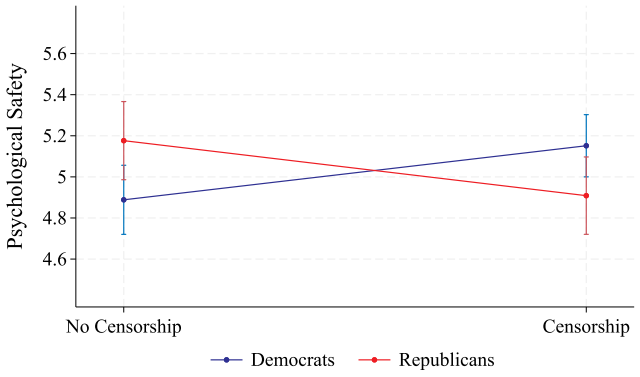
Results

Table 1 reports descriptive statistics and correlations for Study 2. We used linear regression with an interaction term between organizational censorship and partisan affiliation to test Hypothesis 1. As expected, partisan affiliation moderated the association between organizational censorship and psychological safety ($\beta = -0.53$, $p < 0.01$), such that the association was positive for Democrats (slope = 0.27, $p < 0.05$) but negative for Republicans (slope = −0.26, $p < 0.05$). See Figure 1. In short, in organizations that do versus do not censor prejudiced posts, Democratic employees tend to feel psychologically safer but Republican employees tend to feel less psychologically safe.

Study 3

Given that our findings in Study 2 were observational, we next assessed the causal influence of organizational censorship on psychological safety for Democrats versus Republicans (Hypothesis 1) in an experimental context

Figure 1. (Color online) Study 2 Censorship and Psychological Safety



Notes. The figure presents the association between organizational censorship and psychological safety for Democrats versus Republicans in Study 2 with 95% confidence intervals (CIs). Note that the interaction is statistically significant ($p = 0.003$) despite the overlapping CIs (see Knol et al. 2011).

(Study 3). We recruited 704 participants to complete a survey experiment in which they were asked to imagine that a coworker posted a prejudiced statement on a popular social media platform and were randomly assigned to imagine that the company they worked for either censored or did not censor that coworker. Participants then reported their own expected psychological safety.

Measures

Psychological Safety. We used the same seven-item index from Study 2 ($\alpha = 0.92$).

Organizational Censorship. We coded whether participants were randomized to imagine that, in response to the prejudiced post, their organization censored their coworker (= 1) or did nothing (= 0).

Partisan Affiliation. We coded *partisan affiliation* in the same manner as in Study 1.

Results

Table 2 reports descriptive statistics and correlations for Study 3. We used linear regression with an interaction term between organizational censorship and partisan affiliation to test Hypothesis 1. Again, partisan

Table 1. Study 2 Descriptive Statistics and Correlations

Variable	M	SD	1	2	3	4	5	6
1. Psychological safety	5.04	1.16	1.00					
2. Organizational censorship	0.53	0.50	0.02	1.00				
3. Partisan affiliation	0.41	0.49	0.01	−0.04	1.00			
4. Age	38.91	10.05	0.03	−0.04	0.05	1.00		
5. Gender	0.50	0.50	0.05	0.06	0.12*	−0.07	1.00	
6. Race	0.05	0.23	−0.01	0.05	−0.01	−0.06	0.01	1.00

Notes. "M" stands for mean. $N = 704$. Gender is coded 1 = *Male*, 0 = *Female*. Race is coded 1 = *White* or *non-Latino*, 0 = *Other*.
*$p < 0.05$.

Table 2. Study 3 Descriptive Statistics and Correlations

Variable	M	SD	1	2	3	4	5	6
1. Psychological safety	4.35	1.33	1.00					
2. Organizational censorship[a]	0.50	0.50	0.02	1.00				
3. Partisan affiliation	0.47	0.50	0.02	0.02	1.00			
4. Age	38.50	10.44	0.02	0.00	0.04	1.00		
5. Gender	0.50	0.50	0.05	0.07	0.12*	−0.10*	1.00	
6. Race	0.69	0.46	0.07	0.01	0.32*	0.20*	0.05	1.00

Notes. $N = 704$. Gender is coded 1 = *Male*, 0 = *Female*. Race is coded 1 = *White* or *non-Latino*, 0 = *Other*.
[a]Variable was randomly assigned.
*$p < 0.05$.

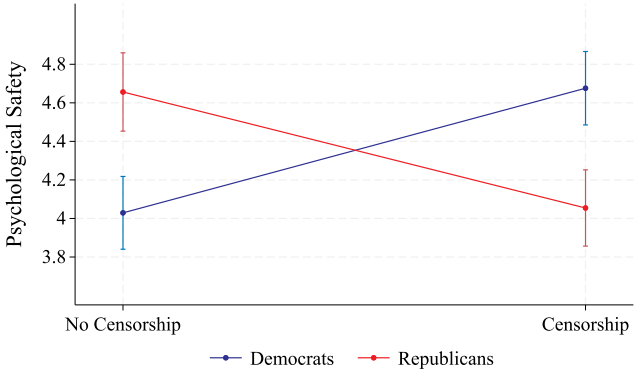
affiliation moderated the association between organizational censorship and psychological safety ($\beta = -1.25$, $p < 0.001$), such that censorship had a positive effect on psychological safety for Democrats (slope = 0.65, $p < 0.001$) but a negative effect on psychological safety for Republicans (slope = −0.60, $p < 0.001$). See Figure 2. Accordingly, organizational censorship of prejudiced social media posts appears to have antithetical causal effects on psychological safety for Democratic versus Republican employees. Note that we replicated these Study 3 findings using Edmondson's (1999) scale without replacing any items and including a manipulation check; see the Online Appendix for details.

Study 4

In Study 4, we assessed the causal influence of specific organizational censorship policies on psychological safety for Democrats versus Republicans (Hypotheses 2a–d) in a conjoint survey experiment. Conjoint designs present participants with one or more hypothetical "profiles" describing a scenario, person, or product (Bansak et al. 2021). Each profile contains a set of attributes (or factors) that might affect the participants' evaluations of the profile. Each attribute consists of multiple levels (or conditions) that are randomly varied to form numerous possible combinations for each profile. Participants evaluate each profile independently. The resulting evaluations are then aggregated to identify participants' multidimensional reactions. Conjoint designs are uniquely valuable because they allow researchers to estimate many causal effects and interactions in a single randomized experiment (Bansak et al. 2021). Conjoints are well suited for the current research because censorship practices are multifaceted, and many factors may differentially affect employees' psychological safety. Additionally, conjoint designs vitiate social desirability bias (Horiuchi et al. 2022), are robust to experimenter demand effects (Mummolo and Peterson 2019), and—in contrast to traditional experiments—have strong *external* validity (Hainmueller et al. 2015, Auerbach and Thachil 2018). Conjoints have been featured in many prominent journals (e.g., *Science*, *Nature*, *Proceedings of the National Academy of Sciences*, *American Sociological Review*, and the *American Political Science Review* (Bansak et al. 2016, Amengual and Bartley 2022, Bor et al. 2023, Kozyreva et al. 2023, Pradel et al. 2024)). But, despite their value, they are underutilized in management; a search of *Organization Science*, *Administrative Science Quarterly*, *Academy of Management Journal*, and the *Journal of Applied Psychology* did not yield any articles with the term "conjoint experiment."

We utilized a single-profile conjoint design in which participants were asked to imagine that a coworker at their company posted a statement on a popular social media platform. Each profile included five randomly assigned attributes: the values emphasized by the organization, the group targeted in the coworker's post, the severity of the statement in the post, whether the coworker's social media account referenced the organization, and how the organization responded to the post (see Figure 3 for the possible levels of each attribute). This design yielded 162 possible unique profiles (a $3 \times 3 \times 3 \times 2 \times 3$ design). We recruited 1,570 participants, and each participant evaluated three different profiles, for a total of 4,710 observations. Note that such a small increase in the number of profiles evaluated (from one in a traditional experiment to three here) does not significantly increase cognitive strain (Bansak et al. 2018, 2021).

Figure 2. (Color online) Study 3 Censorship and Psychological Safety



Note. The figure presents the effects of organizational censorship on psychological safety for Democrats versus Republicans in Study 3 with 95% CIs.

Figure 3. (Color online) Study 4 Example Profile and Randomized Attribute Levels

Example Profile		Randomized Levels	
Attribute	Text with Randomly Selected Level	*N*	Levels
Organizational Values	Imagine that you work for a company that emphasizes its commitment to the free exchange of ideas.	3	[None] [blank]; [Free Speech] that emphasizes its commitment to the free exchange of ideas; [Anti-Prejudice] that emphasizes its commitment to preventing prejudice from spreading
Targeted Group	One day, you see that one of your coworkers posted the following statement on a popular social media platform: "Republicans	3	[In-Party] Republicans/Democrats; [Transgender] Trans people; [Blacks] Blacks
Speech Severity	have too much influence in our society."	3	[Criticism] have too much influence in our society; [Insult] are all just plain stupid; [Threat] should be rounded up and punched in the face
Affiliation Referenced	You also noticed that the coworker references working at your company on their social media profile.	2	[No Affiliation] does not reference; [Affiliation] references
Organizational Censorship	In response to this post, your company did nothing.	3	[No Censorship] did nothing; [Censorship] required your coworker to remove the post; fired your coworker

Notes. Red indicates explanatory text not shown to participants. Blue indicates text that randomly varied across profiles. Randomized levels are separated by semicolons. The figure presents the conjoint design as a table to clarify the different attributes. However, participants were shown the attributes in paragraph form.

First, we compared companies that emphasized a commitment to the free exchange of ideas versus companies that emphasized preventing prejudice from spreading. These organizational values reflect the conflicting values at stake in censorship debates. We also included a blank level for this attribute because many companies do not explicitly endorse one of these values over the other.

Next, we compared coworker prejudiced posts that targeted different social groups. We included a level for posts targeting the participant's political party to create a baseline comparison group to which each participant belonged and which is not marginalized in society. We also included levels for two marginalized groups: one that is traditionally acknowledged as marginalized in society with a long history of social and legal protections from discrimination (Blacks) and one that is not (transgender people).

We then compared the severity of the prejudiced statements. We included levels for mere criticism ("[group members] have too much influence in society") versus an insult ("[group members] are just plain stupid") versus a threat ("[group members] should all be rounded up and punched in the face").

We also compared posts from coworkers whose social media accounts referenced versus did not reference the organization for which they work in order to test whether these forms of integrating versus segmenting behavior on the part of the coworker influenced reactions to organizational censorship.

Finally, we compared situations in which the participant's hypothetical organization responded to the coworker's prejudiced post by doing nothing or censoring the coworker (by either requiring the coworker to remove the post or firing the coworker) in order to test whether these forms of boundary control on the part of the organization influenced reactions to organizational censorship.

Like most conjoint survey experiments (Bansak et al. 2016, Nalick et al. 2020, Amengual and Bartley 2022), we did not include manipulation checks because asking about one randomized attribute would reveal the attribute of interest to participants, which could affect responses in subsequent profiles. But asking about all five attributes after each of the three profiles would substantially prolong the survey.

Attributes and Measures

Organizational Censorship. The results of four pilot studies indicated that participants did not distinguish between organizations censoring a coworker by requiring them to remove the post versus firing them. Accordingly, we first evaluated whether to pool these two levels of censorship. As expected, we could not reject the null hypothesis of no difference in means between these levels ($t(3,144) = 1.28$, $M_{diff} = 0.08$, $SE = 0.06$, $d = 0.05$, $p = 0.20$). Therefore, consistent with our preregistration, we pooled these levels and coded organizational censorship as a dichotomous variable (1 = any censorship; 0 = *did nothing*). We did not use the term "censor" to avoid prescriptive responses due to partisan cues.

Attribute Variables. We used dichotomous indicators for each randomized level of the other attributes (e.g., we used indicators for the "In-Party [Republicans/Democrats]," "Blacks," and "Trans people" levels of the *Targeted Group* attribute) to assess their effects on psychological safety (Figure 3).

Psychological Safety. After each profile, participants reported the level of psychological safety they would

feel having seen their company's response to their coworker posting the prejudiced statement using our preregistered, original three-item index, which draws on extant definitions of the construct ($\alpha = 0.92$ (Kahn 1990, Edmondson 1999)). We adopted this shortened scale in the conjoint study to avoid asking participants to repeatedly complete the longer scale after reading each of three separate scenarios.

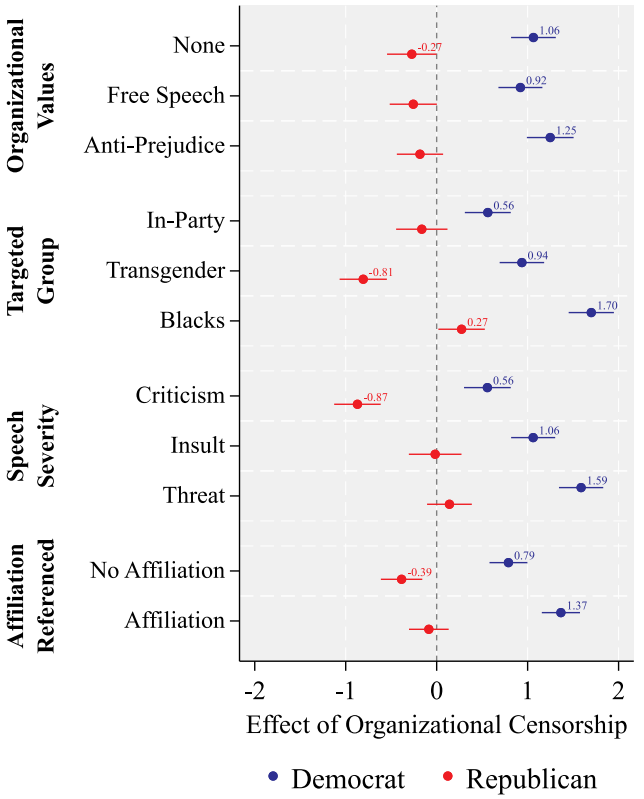
Partisan Affiliation. We coded *partisan affiliation* in the same manner as in Study 1.

Results

Table 3 reports descriptive statistics and correlations for Study 4. We used linear regression with an interaction term between organizational censorship and each level of the other four attributes to test Hypotheses 2a–d. We clustered standard errors on participant to account for possible interdependencies in a participant's ratings. Figure 4 presents the average marginal component effects (AMCEs) of organizational censorship at each level of the other attributes. This figure allows us to examine the effect of organizational censorship on psychological safety under different conditions. For instance, the top line of the figure indicates that, in an organization that does not explicitly endorse a free speech or antiprejudice value and censors a coworker (versus doing nothing), Democrats feel psychologically safer (AMCE = 1.06, $p < 0.00$) but Republicans feel less psychologically safe (AMCE = −0.27, $p = 0.04$). Notice that, in general, organizational censorship increases psychological safety for Democrats and decreases psychological safety for Republicans across the different levels of each attribute. However, as we theorized for Hypothesis 2, there may be some forms of organizational censorship that, by creating an equilibrium, maximize psychological safety by making Democrats feel psychologically safer without making Republicans feel less psychologically safe. And, indeed, our results indicate that several situations meet these optimal criteria for organizational censorship.

First, regarding organizational values, we found that organizational censorship decreased psychological safety for Republicans when the organization either

Figure 4. (Color online) Study 4 Censorship and Psychological Safety



Notes. The figure presents the effects of organizational censorship on psychological safety for Democrats versus Republicans at each level of each attribute in Study 4. Horizontal lines indicate 95% CIs.

Table 3. Study 4 Descriptive Statistics and Correlations

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Psychological safety	4.18	1.70	1.00												
2. Free speech values[a]	0.34	0.47	0.01	1.00											
3. Antiprejudice values[a]	0.34	0.47	0.01	−0.51*	1.00										
4. Transgender targeted[a]	0.33	0.47	0.04*	−0.01	0.01	1.00									
5. Blacks targeted[a]	0.33	0.47	0.02	0.00	0.01	−0.50*	1.00								
6. Insult severity[a]	0.34	0.47	−0.00	−0.01	−0.01	0.00	−0.01	1.00							
7. Threat severity[a]	0.34	0.47	0.00	−0.01	0.00	0.00	−0.01	−0.51*	1.00						
8. Affiliation referenced[a]	0.50	0.50	0.02	0.01	0.00	0.02	0.00	−0.03*	0.01	1.00					
9. Organizational censorship[a]	0.67	0.47	0.12*	−0.00	0.02	−0.03*	−0.00	−0.00	0.01	0.02	1.00				
10. Partisan affiliation	0.47	0.50	−0.02	−0.00	0.03*	0.00	−0.01	−0.00	0.01	0.00	−0.01	1.00			
11. Age	39.35	10.89	−0.01	0.01	0.01	0.00	−0.00	0.00	−0.00	−0.02	−0.01	0.10*	1.00		
12. Gender	0.51	0.50	0.07*	−0.01	−0.01	−0.02	0.02	0.02	0.00	−0.01	0.01	0.05*	−0.12*	1.00	
13. Race	0.67	0.47	0.02	0.02	−0.01	0.02	−0.02	−0.01	0.00	0.00	0.01	0.28*	0.25*	0.03*	1.00

Notes. $N = 4{,}710$ observations (1,570 participants). Gender is coded 1 = *Male*, 0 = *Female*. Race is coded 1 = *White* or *non-Latino*, 0 = *Other*.
[a]Variable was randomly assigned.
*$p < 0.05$.

emphasized its commitment to the free exchange of ideas (AMCE = −0.26, $p = 0.052$) or did not emphasize any particular value (AMCE = −0.27, $p = 0.048$). In contrast, consistent with Hypothesis 2a, this negative effect was not statistically significant for Republicans when the organization emphasized its commitment to preventing prejudice from spreading. In other words, although, on average, censorship makes Republicans feel less psychologically safe, this negative effect does not emerge when organizations censor employees in a manner that is consistent with its values. Censorship always increased psychological safety for Democrats, regardless of its organizational values or lack thereof. But, also consistent with Hypothesis 2a, this effect was strongest when an organization emphasized antiprejudice values (AMCE = 1.25, $p < 0.00$). These results suggest that both Republicans and Democrats are most comfortable when the organization's censorship practice is consistent with its endorsed values.

Next, regarding the targeted group, we found that censorship decreased psychological safety for Republicans when organizations censored posts targeting transgender people (AMCE = −0.81, $p < 0.00$). In contrast, consistent with Hypothesis 2b, this negative effect did not emerge for Republicans when the organization censored posts targeting Blacks; in fact, Republicans actually felt safer when an organization censored such posts (AMCE = 0.27, $p = 0.04$). Organizational censorship always increased psychological safety for Democrats, regardless of which social group was targeted. But, also consistent with Hypothesis 2b, organizational censorship had the strongest positive effect on Democrats' psychological safety when it censored posts targeting Blacks (AMCE = 1.70, $p < 0.00$). These findings suggest that both Republicans and Democrats are more comfortable with organizational censorship that protects traditionally marginalized groups.

Regarding speech severity, for Republicans we found that organizational censorship of posts containing mere criticism decreased psychological safety (AMCE = −0.87, $p < 0.00$), whereas censorship of posts containing insulting or (consistent with Hypothesis 2c) threatening language did not have significant negative effects. Organizational censorship always increased psychological safety for Democrats, regardless of the statement's severity. Also consistent with Hypothesis 2c, censorship of threatening posts had the strongest positive effect on Democrats' psychological safety (AMCE = 1.59, $p < 0.00$). Thus, both Republicans and Democrats are more comfortable with censoring threatening, compared with merely critical, prejudiced statements, indicating that the severity of censored speech also matters for reactions to censorship.

Finally, consistent with Hypothesis 2d, organizational censorship of a coworker's post only decreased psychological safety for Republicans when the coworker's social media account did not reference the organization (AMCE = −0.39, $p < 0.00$). Censorship always increased psychological safety for Democrats, regardless of references to the company. Yet, also consistent with Hypothesis 2d, censorship had a stronger positive effect on psychological safety for Democrats when the coworker's account referenced the organization (AMCE = 1.37, $p < 0.00$). Thus, both Republicans and Democrats are more comfortable with censorship of prejudiced posts when employees' social media accounts reference their organization.

Altogether, these results indicate that organizations can increase psychological safety for Democrats *without* undermining psychological safety for Republicans by (a) emphasizing organizational values that are consistent with its censorship practices, (b) censoring social media posts that target Blacks (and perhaps other groups traditionally and broadly recognized as marginalized in society), (c) censoring posts that contain threatening language, and (d) only censoring posts if an employee's social media account references their affiliation with the organization.

Discussion

Freedom of speech receives broad support worldwide (Wike and Schumacher 2020), yet the proliferation of prejudiced speech on social media has sparked intense ideological debates about censorship in society. As a corollary in the workplace, we documented in a U.S. nationally representative prestudy that Democrats and Republicans differ in their concerns regarding coworkers' prejudiced social media posts and being censored by their employers for allegedly prejudiced posts. We then explored employees' reactions to organizational censorship in open-ended, qualitative responses (Study 1). And, in observational and experimental surveys, we found that Democrats feel psychologically safer when their organization censors prejudiced posts whereas such control undermines Republicans' sense of psychological safety (Hypothesis 1; Studies 2 and 3). Thus, we exposed a moral dilemma in which organizations must determine whose psychological safety to prioritize but also theorized they could craft policies that balance safeguards against free expression. In a conjoint experiment, we found that organizations can successfully manage the differential psychological safety of their employees when (Hypothesis 2a) organizational values align with censorship practices, and the organization censors (Hypothesis 2b) anti-Black speech, (Hypothesis 2c) threatening speech, and (Hypothesis 2d) prejudiced posts from accounts that reference one's organization. As such, organizations can improve psychological safety for liberals without impinging on psychological safety for conservatives by implementing social media policies that safeguard against prejudice

and preserve an adequate degree of employee "free" speech. Altogether, our multimethod package of studies balanced open-ended exploration of a new social phenomenon, external validity, and the benefits of experimental designs, which facilitate causal inferences, internal validity, and reproducible research.

Theoretical Contributions

The ideological tension between safeguards and free expression that we have identified is a timely instantiation of the more general push-pull dynamic between control and autonomy in organizations (Putnam et al. 2016). Beyond documenting this phenomenon, we make several theoretical contributions to the psychological safety literature. Scholars have largely studied psychological safety as a mechanism, often moderated by other factors (Newman et al. 2017, Edmondson and Bransby 2023). Here, we join extant research on what gives rise to psychological safety (Edmondson and Lei 2014, Newman et al. 2017), as well as the boundary conditions of its antecedents (Harvey et al. 2019, Jiang et al. 2019, Dwivedi et al. 2023). Our cross-level approach highlights that the same organizational practice can enhance or diminish psychological safety for different employees. This tension raises the theoretical question of whether psychological safety can be effectively managed by taking an either/or approach. Enacting censorship to enhance psychological safety for liberals would come at a cost for conservatives. Refraining from censorship to enhance psychological safety for conservatives would come at a cost for liberals. As such, we also build on the few psychological safety studies that have considered employees' demographics (e.g., Singh et al. 2013, Dwivedi et al. 2023) by drawing attention to political ideology as a relevant, understudied social identity marker that moderates the effects of an organization-level policy.

Our research highlights that efforts to promote psychological safety can simultaneously subvert this goal. Accordingly, we also examined how to constructively manage tension that stems from competing stakeholder demands. We theorized that creating an equilibrium, in this case through cross-party consensus, would be more effective than taking an either/or approach (Schad et al. 2016). We believe that research on cultivating psychological safety in today's pluralistic organizations can generate theoretical value by investigating less conventional and more nuanced ways to maximize psychological safety for all employees. In our work, because censorship policies do not have uniform effects on Democrats and Republicans, the best way to promote psychological safety was to increase psychological safety for Democrats without undermining it for Republicans. Specifically, striking this equilibrium first involved censorship congruent with an organization's values and the broader social unacceptability of anti-Black speech. Balance was then created by allowing for nonthreatening speech or any speech not affiliated with one's organization. Indeed, "the axis between freedom and control is not ideally managed by increasing one at the expense of the other" (Graetz and Smith 2009, p. 23), and both sides of a tension are "necessary but not sufficient for the well-being of the organization" (Ashforth and Reingen 2014, p. 475). Thus, this work advances our understanding of how to manage psychological safety when antecedents, such as organizational censorship, yield divergent effects on employees. Here, we have theorized and found four ways to enhance psychological safety for liberals vis-à-vis safeguards without compromising psychological safety for conservatives vis-à-vis free expression.

Of equal importance, we contribute to the psychological safety literature by integrating research on boundary management (Rothbard and Ollier-Malaterre 2016, Behrend et al. 2024), including (mis)alignment between individual preferences and organizational policies (Rothbard et al. 2005, Ramarajan and Reid 2013). Quite notably, prior work has not considered how psychological safety is shaped by boundary blurring or organizational control across the online personal/professional boundary (Edmondson and Bransby 2023). Organizations exert boundary control to *positively* influence work-specific *behaviors* (Perlow 1998, Stanko and Beckman 2015), but even policies intended to support employees' nonwork lives can evoke negative attitudes (Flynn and Leslie 2023). Building on such research, we draw attention to the positive *and* negative effects of online censorship on psychological safety, an affective state, depending on one's politics. In doing so, we also highlight that, although an employee's sense of psychological safety pertains to the work domain (Edmondson 1999), psychological (un)safety can develop in "personal," nonwork contexts. As such, we extend theorizing about fostering psychological safety from antecedents that are more clearly work related (e.g., leadership behaviors (Coutifaris and Grant 2021)) to those that are not—here, censorship of prejudiced posts on personal social media accounts.

Finally, we contribute to the growing body of organizational research on societal politics in the workplace. Some studies have examined ideological influence on strategic decisions (e.g., social political activism (Wowak and Busenbark 2024)) and employment decisions (see Swigart et al. 2020), as well as how ideological diversity shapes group decision making (Solomon and Hall 2023). Most notably, prior research that has incorporated social media has shown that hireability judgments are influenced by perceived political similarity to job candidates gleaned through cybervetting (Roth et al. 2020, Wade et al. 2020) and doxing (Roth et al. 2024). Whereas that research emphasizes

in-group/out-group dynamics, we identify asymmetric partisan reactions to censorship policies among existing organizational members. Of course, organizations are unlikely to achieve political homogeneity despite biased hiring. But, more importantly, political polarization is not limited to (dis)similarity and in-party and out-party attitudes (e.g., liking and suspicion (Rosen et al. 2024)). Indeed, in our work, increased and decreased psychological safety does *not* reflect reactions to in- versus out-party coworkers, but rather differential ideological reactions to a policy, which is less immediately apparent and may represent a more novel theoretical contribution. Our work also dovetails a broader discourse on how organizational factors influence ideologically driven cognitions and behaviors, such as perceptions of ideological polarization in response to monitoring (Patil and Bernstein 2022) and the reduced likelihood of organizational exit following workplace protesting in environments more receptive to certain social movements (Rheinhardt et al. 2024).

Practical Implications

This research also offers valuable practical implications regarding psychological safety, "free" speech, and censorship. Given the importance of psychological safety in diverse groups and organizations (Singh et al. 2013, Leigh and Melwani 2022), our work suggests questioning policies predicated on the premise that psychological safety is cultivated in the same manner for all employees. Identifying the ideological tension regarding censorship and psychological safety was important because identification is critical to managing tension in practice and "capturing its enlightening potential" (Lewis 2000, p. 763). Our work highlights that political polarization is so strong, as also evidenced in research on politically biased hiring (Roth et al. 2020), that the same well-intended censorship policy can have impacts that are diametrically opposed and unintended for some employees.

Our research also provides theory-based rather than purely phenomenon-based suggestions for how the safeguards/free speech tension may be managed effectively. Tensions help managers recognize that in diverse, contemporary organizations, decision making is interdependent (Lewis 2000, Carlos and Lewis 2018). In our work, this reality means that promoting safeguards inevitably affects free speech and vice versa; thus, censorship decisions must be made with consequences of both in mind. Our findings indicate that deliberately adopting both elements of the tension is necessary to prevent censorship policies from backfiring, just as diversity, equity, and inclusion (DEI) initiatives should be implemented more deliberately to avoid backlash (Burnett and Aguinis 2024). Indeed, we found evidence that four factors moderate the effects of censorship on psychological safety in a manner that benefits Democrats without negatively impacting Republicans. Thus, from a managerial perspective, striking an equilibrium between safeguards and free expression appears more effective than an either/or approach.

Indeed, our findings suggest that organizations can promote psychological safety for liberals without undermining psychological safety for conservatives by crafting social media policies narrowly to target threatening speech and speech that targets traditionally protected groups. And, quite notably, the equilibrium that emerged in our work provides two practical escape valves for employee free speech. Specifically, one can circumvent censorship if they can refrain from threatening speech (and stick to merely critical or insulting speech) or refrain from referencing their organization on their social media account. Although the degree of digital cultural capital that is generally required to successfully navigate the personal/professional online boundary is significant (Ollier-Malaterre et al. 2019), we believe that opting into the free speech avenues that we have identified should be relatively feasible for most employees. And from the standpoint of an employee who ultimately prefers to maximize autonomy, our findings also suggest that maintaining one's freedom of expression may take the form of self-selecting into an organization that both values *and* prioritizes free speech over preventing prejudice from spreading, or self-selecting into an organization that rejects the Overton window in favor of its own standards for acceptable speech that one also personally endorses. Self-selection may be feasible given the variety of censorship positions organizations take, though some organizations shift positions over time. For example, "it is now permitted to call gay people mentally ill on Facebook, Threads and Instagram," according to Meta's recent relaxation of rules around hate speech (Ortutay 2025). It is unclear whether such changes by social media companies will affect other organizations' tolerance of employee speech.

We believe our results also offer value for both practitioners and organizational members. Liberals must accept areas of cross-party consensus to achieve an adequate balance of psychological safety for themselves and their political rivals. Conservatives must accept the legal reach of organizational control even if it threatens their psychological safety. The blurriness between public and private life on social media has rendered "free speech" quite complicated (McFarland and Ployhart 2015, Behrend et al. 2024). Having integrated political science insights into academic discourse about employee speech online, we hope our work is leveraged by stakeholders crafting censorship policies amid the free speech debate.

Limitations and Future Directions

We now turn to limitations and future directions. First, our prestudy was exploratory and simply provided a phenomenological basis for developing our theory. Study 1 was also exploratory and not intended to serve as a comprehensive qualitative design. Studies 2 and 3 directly tested Hypothesis 1 but relied on recall and a hypothetical scenario, respectively. We also note that variations of the term "censor" (Studies 1–3) could have prompted prescriptive partisan reactions and responses. To alleviate this limitation, Study 4 avoided labeling censorship as such. Because Study 4 was a conjoint survey experiment, we adapted and shortened Edmondson's (1999) commonly used seven-item scale, as shorter scales are preferred in such designs. Although that may be a limitation, our three-item scale had high reliability (Studies 2–4) and a strong correlation with the seven-item scale (Studies 2–3); moreover, results were robust to either. And, of course, because we were unable to randomly assign participants to work in organizations with different censorship policies, we used hypothetical scenarios in Studies 3–4 (cf. Levine et al. 2023).

More substantively, Study 4 randomized three targeted groups (Blacks, transgender people, and in-party members) to examine reactions to censorship of prejudiced speech. Each has significant influence in today's social discourse. But examining other targeted groups would enhance generalizability regarding marginalized versus nonmarginalized groups and marginalized groups that have received more versus fewer protections in society. Also, psychological distances to the targets of censored posts and to censored coworkers are potential boundary conditions that tap into social identity. For instance, the positive effects for Democrats may disappear when a close colleague is censored, presumably for speech that they endorse (e.g., a pro-Gaza post). And we did not examine posts targeting the outparty, such as a Democrat's reaction to censorship of a post targeting Donald Trump, which could undermine the psychological safety of Democratic employees because of perceived misalignment of values. Republican participants in Study 4 might have felt psychologically safer if the organization censored anti-White or transgender rights speech. However, we note that Studies 1–3 refer to generic "prejudiced" speech, and Republicans were still more opposed to and felt less psychologically safe with censorship. We also note that Democrats felt safer in response to censorship of posts targeting the in-party whereas such censorship did not affect Republicans.

More generally, studies should also examine how employees learn about organizational censorship. Compared with knowledge of their organization's policy, awareness of specific instances of censorship could exacerbate partisan reactions. Alternatively, such awareness might complicate the narrative, such that liberals feel less psychologically safe knowing more definitively that a coworker harbors prejudiced views that warranted censorship and conservatives feel more psychologically safe seeing that censorship was warranted. Additionally, awareness may come from different sources, including from gossip, from the organization through termination or a policy enforcement notice, or from the censored coworker through direct conversation, indirect communication on social media, or observations of the coworker's changed behavior on social media. Although greater psychological distance to the source is unlikely to alter our effects, as censorship policies are already abstract, greater proximity may make censorship more concrete and amplify partisan effects.

Next, to enhance generalizability, we did not recruit from a single organization. But future work might consider how the extent and type of diversity in an organization influence psychological safety. For example, if an organization is especially politically heterogenous/homogeneous, then the ideological tension we uncovered should be more/less problematic for fostering a psychologically safe environment. Studies may also recruit enough Black and transgender participants to conduct analyses exploring in-group moderators beyond party. We also focused on censorship of prejudiced speech, but examining a range of social media content would indicate whether differential effects consistently emerge for liberals and conservatives. Organizational reach regarding online prejudice could be uniquely polarizing.

Finally, this research was conducted in a politically polarized time in American society and in a presidential election year. Our results may not hold in less polarized countries. And, although many countries are also polarized (Silver et al. 2021), freedom of expression is more popular in the United States than in most other regions. Thus, other cultures may be less divided in response to organizational censorship. Also, events such as a protest or judicial decision that politicize free speech and render censorship debates more salient may moderate our effects. Studying censorship over time could capture these events and changes in the acceptability of speech targeting various social groups—both of which should enhance knowledge of psychological safety's stability and how employees' differential reactions evolve. Indeed, liberals' and conservatives' censorship attitudes have realigned in the past (Chong et al. 2024).

In sum, our research uncovered an ideological tension that pits safeguards against free expression and demonstrated the value of embracing tension versus an either/or approach. As societal polarization increasingly seeps into the workplace, maintaining the social-relational fabric of organizations will require crafting policies that balance ideological demands and

considering how control over nonwork factors may cultivate or impinge on psychological safety for some stakeholders versus others.

References

Al-Kire RL, Miller CA, Pasek MH, Perry SL, Wilkins CL (2024) White by another name? Can anti-Christian bias claims serve as a racial dog whistle? *Psych. Sci.* 35(4):415–434.

Amengual M, Bartley T (2022) Global markets, corporate assurances, and the legitimacy of state intervention: Perceptions of distant labor and environmental problems. *Amer. Sociol. Rev.* 87(3):383–414.

Amira K, Wright JC, Goya-Tocchetto D (2021) In-group love versus out-group hate: Which is more important to partisans and when? *Political Behav.* 43(2):473–494.

Ashforth BE, Reingen PH (2014) Functions of dysfunction: Managing the dynamics of an organizational duality in a natural food cooperative. *Admin. Sci. Quart.* 59(3):474–516.

Ashforth BE, Kreiner GE, Fugate M (2000) All in a day's work: Boundaries and micro role transitions. *Acad. Management Rev.* 25(3):472–491.

Astor M (2019) How the politically unthinkable can become mainstream. *New York Times* (February 26), https://www.nytimes.com/2019/02/26/us/politics/overton-window-democrats.html.

Auerbach AM, Thachil T (2018) How clients select brokers: Competition and choice in India's slums. *Amer. Political Sci. Rev.* 112(4): 775–791.

Babu N, De Roeck K, Raineri N (2020) Hypocritical organizations: Implications for employee social responsibility. *J. Bus. Res.* 114:376–384.

Banghart S, Etter M, Stohl C (2018) Organizational boundary regulation through social media policies. *Management Comm. Quart.* 32(3):337–373.

Bansak K, Hainmueller J, Hangartner D (2016) How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science* 354(6309):217–222.

Bansak K, Hainmueller J, Hopkins DJ, Yamamoto T (2018) The number of choice tasks and survey satisficing in conjoint experiments. *Political Anal.* 26(1):112–119.

Bansak K, Hainmueller J, Hopkins DJ, Yamamoto T (2021) Conjoint survey experiments. Green DP, Druckman JN, eds. *Advances in Experimental Political Science* (Cambridge University Press, Cambridge, UK), 19–41.

Barden J, Rucker DD, Petty RE (2005) "Saying one thing and doing another": Examining the impact of event order on hypocrisy judgments of others. *Personality Soc. Psych. Bull.* 31(11):1463–1474.

Behrend TS, Ravid DM, Thapa S (2024) Implications of social media for a changing work landscape. *Annual Rev. Organ. Psych. Organ. Behav.* 11(1):337–361.

Bilewicz M, Soral W, Marchlewska M, Winiewski M (2017) When authoritarians confront prejudice. Differential effects of SDO and RWA on support for hate-speech prohibition. *Political Psych.* 38(1):87–99.

Boone R (2023) Experts say attacks on free speech are rising across the US. *AP News*. Retrieved March 10, 2024, https://apnews.com/article/first-amendment-free-speech-censorship-mccarthyism-815865bafa52bb821400be15fdf76119.

Bor A, Jørgensen F, Petersen MB (2023) Discriminatory attitudes against unvaccinated people during the pandemic. *Nature* 613(7945): 704–711.

Brehm JW (1966) *A Theory of Psychological Reactance* (Academic Press, New York).

Brehm SS, Brehm JW (2013) *Psychological Reactance: A Theory of Freedom and Control* (Academic Press, New York).

Brenan M (2024) Same-sex relations, marriage still supported by most in U.S. *Gallup.com*. Retrieved September 1, https://news.gallup.com/poll/646202/sex-relations-marriage-supported.aspx.

Burnett L, Aguinis H (2024) How to prevent and minimize DEI backfire. *Bus. Horizons* 67(2):173–182.

Carlos WC, Lewis BW (2018) Strategic silence: Withholding certification status as a hypocrisy avoidance tactic. *Admin. Sci. Quart.* 63(1):130–169.

Cha SE, Edmondson AC (2006) When values backfire: Leadership, attribution, and disenchantment in a values-driven organization. *Leadership Quart.* 17(1):57–78.

Cho A (2018) Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media Soc.* 20(9): 3183–3200.

Chong D, Citrin J, Levy M (2024) The realignment of political tolerance in the United States. *Perspect. Politics* 22(1):131–152.

Collins NL, Miller LC (1994) Self-disclosure and liking: A meta-analytic review. *Psych. Bull.* 116(3):457–475.

Cortina LM, Kabat-Farr D, Magley VJ, Nelson K (2017) Researching rudeness: The past, present, and future of the science of incivility. *J. Occupational Health Psych.* 22(3):299–313.

Costello M, Hawdon J, Ratliff T, Grantham T (2016) Who views online extremism? Individual attributes leading to exposure. *Comput. Human Behav.* 63:311–320.

Coutifaris CGV, Grant AM (2021) Taking your team behind the curtain: The effects of leader feedback-sharing and feedback-seeking on team psychological safety. *Organ. Sci.* 33(4):1574–1598.

Crandall CS, Stangor C (2005) Conformity and prejudice. Dovidio JF, Glick P, Rudman LA, eds. *On the Nature of Prejudice* (John Wiley & Sons, Ltd, Bognor Regis, UK), 293–309.

Creary SJ, Locke K (2022) Breaking the cycle of overwork and recuperation: Altering somatic engagement across boundaries. *Organ. Sci.* 33(3):873–900.

Daniller A (2021) Majorities of Americans see at least some discrimination against Black, Hispanic and Asian people in the U.S. Pew Research Center, Washington, DC.

Dell Technologies (2023) Global social media policy. Dell Technologies, Round Rock, TX.

Dwivedi P, Gee IH, Withers MC, Boivie S (2023) No reason to leave: The effects of CEO diversity-valuing behavior on psychological safety and turnover for female executives. *J. Appl. Psych.* 108(7): 1262–1276.

Edmondson A (1999) Psychological safety and learning behavior in work teams. *Admin. Sci. Quart.* 44(2):350–383.

Edmondson AC (1996) Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *J. Appl. Behav. Sci.* 32(1): 5–28.

Edmondson AC (2003) Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *J. Management Stud.* 40(6):1419–1452.

Edmondson AC (2018) *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*, 1st ed. (Wiley, Hoboken, NJ).

Edmondson AC, Bransby DP (2023) Psychological safety comes of age: Observed themes in an established literature. *Annual Rev. Organ. Psych. Organ. Behav.* 10(1):55–78.

Edmondson AC, Lei Z (2014) Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Rev. Organ. Psych. Organ. Behav.* 1(1):23–43.

Elliot AJ, Devine PG (1994) On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *J. Personality Soc. Psych.* 67(3):382–394.

Elsamra M (2024) Delta removes an employee, changes its uniform policy after backlash over social media post perceived to be anti-Palestinian. *CNN*. Retrieved November 12, https://www.cnn.com/2024/07/12/us/delta-employee-social-post-palestinian-reaj/index.html.

Festinger L (1962) *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford, CA).

Flynn E, Leslie LM (2023) Progressive or pressuring? The signaling effects of egg freezing coverage and other work–life policies. *J. Appl. Psych.* 108(1):1–26.

Forbes Human Resources Council (2017) Council post: Why your business needs a social media policy and eight things it should cover. *Forbes* (May 25), https://www.forbes.com/sites/forbeshumanresourcescouncil/2017/05/25/why-your-business-needs-a-social-media-policy-and-eight-things-it-should-cover/.

Frazier ML, Fainshmidt S, Klinger RL, Pezeshkan A, Vracheva V (2017) Psychological safety: A meta-analytic review and extension. *Personnel Psych.* 70(1):113–165.

Frey E, Adams GS, Pfeffer J, Belmi P (2023) What we (do not) know about punishment across organizational boundaries. *J. Management* 49(1):196–236.

Gibson K (2023) Growing number of workers lose their jobs for antisemitic social media posts. *CBS News*. Retrieved November 12, 2024, https://www.cbsnews.com/news/citigroup-fires-banker-for-antisemitic-social-media-post/.

Goncalo JA, Chatman JA, Duguid MM, Kennedy JA (2015) Creativity from constraint? How the political correctness norm influences creativity in mixed-sex work groups. *Admin. Sci. Quart.* 60(1):1–30.

Graetz F, Smith ACT (2008) The role of dualities in arbitrating continuity and change in forms of organizing. *Internat. J. Management Rev.* 10(3):265–280.

Graetz F, Smith ACT (2009) Duality theory and organizing forms in change management. *J. Change Management* 9(1):9–25.

Graham J, Meindl P, Koleva S, Iyer R, Johnson KM (2015) When values and behavior conflict: Moral pluralism and intrapersonal moral hypocrisy. *Soc. Personality Psych. Compass* 9(3):158–170.

Graupmann V, Jonas E, Meier E, Hawelka S, Aichhorn M (2012) Reactance, the self, and its group: When threats to freedom come from the ingroup versus the outgroup. *Eur. J. Soc. Psych.* 42(2):164–173.

Graves P (2021) Ask an expert: Online behavior. SHRM, Alexandria, VA.

Gray K, Kubin E (2024) Victimhood: The most powerful force in morality and politics. Gawronski B, ed. *Advances in Experimental Social Psychology*, vol. 70 (Academic Press, New York), 137–220.

Greenbaum RL, Mawritz MB, Piccolo RF (2015) When leaders fail to "walk the talk": Supervisor undermining and perceptions of leader hypocrisy. *J. Management* 41(3):929–956.

Hainmueller J, Hangartner D, Yamamoto T (2015) Validating vignette and conjoint survey experiments against real-world behavior. *Proc. Natl. Acad. Sci. USA* 112(8):2395–2400.

Halbesleben JRB, Leroy H, Dierynck B, Simons T, Savage GT, McCaughey D, Leon MR (2013) Living up to safety values in health care: The effect of leader behavioral integrity on occupational safety. *J. Occupational Health Psych.* 18(4):395–405.

Harvey JF, Johnson KJ, Roloff KS, Edmondson AC (2019) From orientation to behavior: The interplay between learning orientation, open-mindedness, and psychological safety in team learning. *Human Relations* 72(11):1726–1751.

Hawkins CB, Nosek BA (2012) Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personality Soc. Psych. Bull.* 38(11):1437–1452.

Hebl M, Cheng SK, Ng LC (2020) Modern discrimination in organizations. *Annual Rev. Organ. Psych. Organ. Behav.* 7(1):57–82.

Hersel MC, Helmuth CA, Zorn ML, Shropshire C, Ridge JW (2019) The corrective actions organizations pursue following misconduct: A review and research agenda. *Acad. Management Ann.* 13(2):547–585.

Hidy K (2020) Social media policies, corporate censorship and the right to be forgiven: A proposed framework for free expression in an era of employer social media monitoring. *Univ. Pennsylvania J. Bus. Law* 22(2):346–385.

Horiuchi Y, Markovich Z, Yamamoto T (2022) Does conjoint analysis mitigate social desirability bias? *Political Anal.* 30(4):535–549.

Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ (2019) The origins and consequences of affective polarization in the United States. *Annual Rev. Political Sci.* 22(1):129–146.

Jiang Z, Hu X, Wang Z, Jiang X (2019) Knowledge hiding as a barrier to thriving: The mediating role of psychological safety and moderating role of organizational cynicism. *J. Organ. Behav.* 40(7):800–818.

Johnson C (2024) Supreme Court justices appear skeptical of Texas and Florida social media laws. *NPR* (February 26), https://www.npr.org/2024/02/26/1233506273/supreme-court-social-media-laws.

Jost JT (2021) *Left and Right: The Psychological Significance of a Political Distinction* (Oxford University Press, Oxford, UK).

Jost JT, Federico CM, Napier JL (2009) Political ideology: Its structure, functions, and elective affinities. *Annual Rev. Psych.* 60(1):307–337.

Kahn WA (1990) Psychological conditions of personal engagement and disengagement at work. *Acad. Management J.* 33(4):692–724.

Karunakaran A, Orlikowski WJ, Scott SV (2022) Crowd-based accountability: Examining how social media commentary reconfigures organizational accountability. *Organ. Sci.* 33(1):170–193.

Keith BE, Magleby DB, Nelson CJ, Orr E, Westlye MC, Wolfinger RE (1992) *The Myth of the Independent Voter* (University of California Press, Berkeley).

Knight Foundation (2020) *The 100 Million Project: The Untold Story of American Non-voters* (Knight Foundation, Miami).

Knol MJ, Pestman WR, Grobbee DE (2011) The (mis)use of overlap of confidence intervals to assess effect modification. *Eur. J. Epidemiology* 26(4):253–254.

Kossek EE, Lautsch BA (2012) Work–family boundary management styles in organizations: A cross-level model. *Organ. Psych. Rev.* 2(2):152–171.

Kozyreva A, Herzog SM, Lewandowsky S, Hertwig R, Lorenz-Spreen P, Leiser M, Reifler J (2023) Resolving content moderation dilemmas between free speech and harmful misinformation. *Proc. Natl. Acad. Sci. USA* 120(7):e2210666120.

Kraatz MS, Block ES (2008) Organizational implications of institutional pluralism. Greenwood R, Oliver C, Suddaby R, Sahlin K, eds. *The SAGE Handbook of Organizational Institutionalism* (SAGE Publications Ltd, London), 243–275.

Lauriano LA, Coacci T (2023) Losing control: The uncertain management of concealable stigmas when work and social media collide. *Acad. Management J.* 66(1):222–247.

Leigh A, Melwani S (2022) "Am I next?" The spillover effects of mega-threats on avoidant behaviors at work. *Acad. Management J.* 65(3):720–748.

Leroy H, Dierynck B, Anseel F, Simons T, Halbesleben JRB, McCaughey D, Savage GT, Sels L (2012) Behavioral integrity for safety, priority of safety, psychological safety, and patient safety: A team-level study. *J. Appl. Psych.* 97(6):1273–1281.

Levendusky M (2009) *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans* (University of Chicago Press, Chicago).

Levine SS, Schilke O, Kacperczyk O, Zucker LG (2023) Primer for experimental methods in organization theory. *Organ. Sci.* 34(6):1997–2025.

Lewis MW (2000) Exploring paradox: Toward a more comprehensive guide. *Acad. Management Rev.* 25(4):760–776.

Li AH, Noland ES, Monteith MJ (2024) Following prejudiced behavior, confrontation restores local anti-bias social norms. *Personality Soc. Psych. Bull.*, ePub ahead of print March 4, https://doi.org/10.1177/01461672241229006.

Liu W, Song Z, Li X, Liao Z (2017) Why and when leaders' affective states influence employee upward voice. *Acad. Management J.* 60(1):238–263.

Lyon D (2018) *The Culture of Surveillance: Watching as a Way of Life*, 1st ed. (Polity, Cambridge, UK).

Mason L (2018) *Uncivil Agreement: How Politics Became Our Identity* (University of Chicago Press, Chicago).

McFarland LA, Ployhart RE (2015) Social media: A contextual framework to guide research and practice. *J. Appl. Psych.* 100(6): 1653–1677.

Mitchell A, Walker M (2021) More Americans now say government should take steps to restrict false information online than in 2018. Pew Research Center, Washington, DC.

Mönke FW, Lievens F, Hess U, Schäpers P (2024) Politics speak louder than skills: Political similarity effects in hireability judgments in multiparty contexts and the role of political interest. *J. Appl. Psych.* 109(1):1–12.

Mummolo J, Peterson E (2019) Demand effects in survey experiments: An empirical assessment. *Amer. Political Sci. Rev.* 113(2): 517–529.

Nalick M, Kuban S, Hill AD, Ridge JW (2020) Too hot to handle and too valuable to drop: An expanded conceptualization of firms' reactions to exchange partner misconduct. *Acad. Management J.* 63(6):1976–2003.

Neeley TB, Leonardi PM (2018) Enacting knowledge strategy through social media: Passable trust and the paradox of nonwork interactions. *Strategic Management J.* 39(3):922–946.

Newman A, Donohue R, Eva N (2017) Psychological safety: A systematic review of the literature. *Human Resource Management Rev.* 27(3):521–535.

Nippert-Eng C (1996) Calendars and keys: The classification of "home" and "work." *Sociol. Forum* 11(3):563–582.

Norris P (2023) Cancel culture: Myth or reality? *Political Stud.* 71(1):145–174.

Ocasio W, Wohlgezogen F (2010) Attention and control. Sitkin S, Cardinal L, Bijlsma-Frankema KM, eds. *Organizational Control: New Directions in Theory and Research* (Cambridge University Press, Cambridge, UK), 191–221.

Ollier-Malaterre A, Jacobs JA, Rothbard NP (2019) Technology, work, and family: Digital cultural capital and boundary management. *Annual Rev. Sociol.* 45(1):425–447.

Ollier-Malaterre A, Rothbard NP, Berg JM (2013) When worlds collide in cyberspace: How boundary work in online social networks impacts professional relationships. *Acad. Management Rev.* 38(4):645–669.

Olmsted K, Lampe C, Ellison NB (2016) Social media and the workplace. Pew Research Center, Washington, DC.

O'Reilly J, Aquino K (2011) A model of third parties' morally motivated responses to mistreatment in organizations. *Acad. Management Rev.* 36(3):526–543.

Ortutay B (2025) Meta rolls back hate speech rules as Zuckerberg cites "recent elections" as a catalyst. *AP News.* Retrieved March 1, https://apnews.com/article/meta-facebook-hate-speech-trump-immigrant-transgender-41191638cd7c720b950c05f9395a2b49.

Ouchi WG, Maguire MA (1975) Organizational control: Two functions. *Admin. Sci. Quart.* 20(4):559–569.

Paluck EL, Porat R, Clark CS, Green DP (2021) Prejudice reduction: Progress and challenges. *Annual. Rev. Psych.* 72(1):533–560.

Parker K, Horowitz JM, Brown A (2022) Americans' complex views on gender identity and transgender issues. Pew Research Center, Washington, DC.

Patil SV, Bernstein ES (2022) Uncovering the mitigating psychological response to monitoring technologies: Police body cameras not only constrain but also depolarize. *Organ. Sci.* 33(2):541–570.

Perlow LA (1998) Boundary control: The social ordering of work and family time in a high-tech corporation. *Admin. Sci. Quart.* 43(2):328–357.

Pew Research Center (2014) The party of nonvoters. Pew Research Center, Washington, DC.

Pillemer J, Rothbard NP (2018) Friends without benefits: Understanding the dark sides of workplace friendship. *Acad. Management Rev.* 43(4):635–660.

Posner BZ (2010) Another look at the impact of personal and organizational values congruency. *J. Bus. Ethics* 97(4):535–541.

Posner BZ, Kouzes JM, Schmidt WH (1985) Shared values make a difference: An empirical test of corporate culture. *Human Resource Management* 24(3):293–309.

Pradel F, Zilinsky J, Kosmidis S, Theocharis Y (2024) Toxic speech and limited demand for content moderation on social media. *Amer. Political Sci. Rev.* 118(4):1895–1912.

Pratto F, Sidanius J, Stallworth LM, Malle BF (1994) Social dominance orientation: A personality variable predicting social and political attitudes. *J. Personality Soc. Psych.* 67(4): 741–763.

Putnam LL, Fairhurst GT, Banghart S (2016) Contradictions, dialectics, and paradoxes in organizations: A constitutive approach. *Acad. Management Ann.* 10(1):65–171.

Quinn RE, Cameron KS, eds. (1988) *Paradox and Transformation: Toward a Theory of Change in Organization and Management* (Ballinger Publishing Co., Cambridge, MA).

Ramarajan L, Reid E (2013) Shattering the myth of separate worlds: Negotiating nonwork identities at work. *Acad. Management Rev.* 38(4):621–644.

Rego A, Melo AI, Bluhm DJ, Pina e Cunha M, Reis Júnior D (2021) Leader-expressed humility predicting team psychological safety: A personality dynamics lens. *J. Bus. Ethics* 174(3):669–686.

Rheinhardt A, Poskanzer EJ, Briscoe F (2024) The career consequences of workplace protest participation: Theory and evidence from the NFL "Take a knee" movement. *Organ. Sci.* 35(3):888–910.

Roccapriore AY, Pollock TG (2023) I don't need a degree, I've got abs: Influencer warmth and competence, communication mode, and stakeholder engagement on social media. *Acad. Management J.* 66(3):979–1006.

Ronson J (2016) *So You've Been Publicly Shamed* (Riverhead Books, New York).

Rosen CC, Koopman J, Gabriel AS, Lee YE, Ezerins M, Roth PL (2024) Hidden consequences of political discourse at work: How and why ambient political conversations impact employee outcomes. *J. Appl. Psych.* 109(6):795–810.

Rosenberg BD, Siegel JT (2018) A 50-year review of psychological reactance theory: Do not read this article. *Motivation Sci.* 4(4):281–300.

Rosenblum M, Schroeder J, Gino F (2020) Tell it like it is: When politically incorrect language promotes authenticity. *J. Personality Soc. Psych.* 119(1):75–103.

Rosette AS, Carton AM, Bowes-Sperry L, Hewlin PF (2013) Why do racial slurs remain prevalent in the workplace? Integrating theory on intergroup behavior. *Organ. Sci.* 24(5):1402–1421.

Roth PL, Bobko P, Ellingson JE, Thatcher JB, Matthews KD, Goldberg CB (2020) Political affiliation and employment screening decisions: The role of similarity and identification processes. *J. Appl. Psych.* 105(5):472–486.

Roth PL, Bobko P, Shan G, Roth RW, Ferrise E, Thatcher JB (2024) Doxing, political affiliation, and type of information: Effects on suspicion, perceived similarity, and hiring-related judgments. *J. Appl. Psych.* 109(5):730–754.

Rothbard NP, Ollier-Malaterre A (2016) Boundary management. Allen TD, Eby LT, eds. *The Oxford Handbook of Work and Family* (Oxford University Press, Oxford, UK), 109–122.

Rothbard NP, Phillips KW, Dumas TL (2005) Managing multiple roles: Work-family policies and individuals' desires for segmentation. *Organ. Sci.* 16(3):243–258.

Rothbard NP, Ramarajan L, Ollier-Malaterre A, Lee SS (2022) OMG! My boss just friended me: How evaluations of colleagues' disclosure, gender, and rank shape personal/professional boundary blurring online. *Acad. Management J.* 65(1):35–65.

Schad J, Lewis MW, Raisch S, Smith WK (2016) Paradox research in management science: Looking back to move forward. *Acad. Management Ann.* 10(1):5–64.

Schein C, Gray K (2018) The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality Soc. Psych. Rev.* 22(1):32–70.

Schlosser AE (2020) Self-disclosure versus self-presentation on social media. *Current Opinion Psych.* 31:1–6.

Silver L, Fetterolf J, Connaughton A (2021) Diversity and division in advanced economies. Pew Research Center, Washington, DC.

Simons T (2002) Behavioral integrity: The perceived alignment between managers' words and deeds as a research focus. *Organ. Sci.* 13(1):18–35.

Singh B, Winkel DE, Selvarajan TT (2013) Managing diversity at work: Does psychological safety hold the key to racial differences in employee performance? *J. Occupational Organ. Psych.* 86(2):242–263.

Smith WK, Lewis MW (2011) Toward a theory of paradox: A dynamic equilibrium model of organizing. *Acad. Management Rev.* 36(2):381–403.

Solomon BC, Hall MEK (2023) When (non) differences make a difference: The roles of demographic diversity and ideological homogeneity in overcoming ideologically biased decision making. *Organ. Sci.* 34(5):1820–1838.

Solomon BC, Hall MEK, Hemmen A, Druckman JN (2024) Illusory interparty disagreement: Partisans agree on what hate speech to censor but do not know it. *Proc. Natl. Acad. Sci. USA* 121(39):e2402428121.

Stanko TL, Beckman CM (2015) Watching you watching me: Boundary control and capturing attention in the context of ubiquitous technology use. *Acad. Management J.* 58(3):712–738.

St. Aubin C, Liedke J (2023) Most Americans favor restrictions on false information, violent content online. Pew Research Center, Washington, DC.

Stohl C, Etter M, Banghart S, Woo DJ (2017) Social media policies: Implications for contemporary notions of corporate social responsibility. *J. Bus. Ethics* 142(3):413–436.

Swigart KL, Anantharaman A, Williamson JA, Grandey AA (2020) Working while liberal/conservative: A review of political ideology in organizations. *J. Management* 46(6):1063–1091.

Thomas LT, Ganster DC (1995) Impact of family-supportive work variables on work-family conflict and strain: A control perspective. *J. Appl. Psych.* 80(1):6–15.

Thompson JA, Bunderson JS (2003) Violations of principle: Ideological currency in the psychological contract. *Acad. Management Rev.* 28(4):571–586.

Torrez B, Hollie L, Richeson JA, Kraus MW (2024) The misperception of organizational racial progress toward diversity, equity, and inclusion. *Amer. Psych.* 79(4):581–592.

Tu Y, Lu X, Choi JN, Guo W (2019) Ethical leadership and team-level creativity: Mediation of psychological safety climate and moderation of supervisor support for creativity. *J. Bus. Ethics* 159:551–565.

Turner SF, Cardinal LB, Burton RM (2017) Research design for mixed methods: A triangulation-based framework and roadmap. *Organ. Res. Methods* 20(2):243–267.

Villatoro D, Andrighetto G, Brandts J, Nardin LG, Sabater-Mir J, Conte R (2014) The norm-signaling effects of group punishment: Combining agent-based simulation and laboratory experiments. *Soc. Sci. Comput. Rev.* 32(3):334–353.

Wade JT, Roth PL, Thatcher JB, Dinger M (2020) Social media and selection: Political issue similarity, liking, and the moderating effect of social media platform. *MIS Quart.* 44(3): 1301–1357.

Wagner T, Korschun D, Troebs CC (2020) Deconstructing corporate hypocrisy: A delineation of its behavioral, moral, and attributional facets. *J. Bus. Res.* 114:385–394.

Walther JB (2022) Social media and online hate. *Current Opinion Psych.* 45:101298.

White MH, Crandall CS (2017) Freedom of racist speech: Ego and expressive threats. *J. Personality Soc. Psych.* 113(3):413–429.

Wike R, Schumacher S (2020) Democratic rights popular globally but commitment to them not always strong. Pew Research Center, Washington, DC.

Womick J, Goya-Tocchetto D, Ochoa NR, Rebollar C, Kapsaskis K, Pratt S, Payne K, Vaisey S, Gray K (2024) Moral disagreement across politics is explained by different assumptions about who is most vulnerable to harm. Preprint, submitted March 26, https://osf.io/preprints/psyarxiv/qsg7j.

Wowak AJ, Busenbark JR (2024) Why do some conservative CEOs publicly support liberal causes? Organizational ideology, managerial discretion, and CEO sociopolitical activism. *Organ. Sci.* 35(4):1388–1408.

Brittany C. Solomon is the Thomas A. and James J. Bruder assistant professor of administrative leadership in the Mendoza College of Business at the University of Notre Dame. Her research interests include bias detection and suppression, political ideology and sociopolitical spillover, and social perception. She received her PhD in psychology from Washington University in St. Louis.

Alexandra K. Scott is a PhD student in organizational behavior in the Mendoza College of Business at the University of Notre Dame. Her research interests include organizational citizenship behaviors, motivation, gender, and meaning.

Matthew E. K. Hall is the David A. Potenziani Memorial College professor of constitutional studies and director of the Rooney Center for the Study of American Democracy at the University of Notre Dame. His research interests include democratic attitudes, polarization, and American politics. He received his PhD in political science from Yale University.