nature reviews psychology

https://doi.org/10.1038/s44159-025-00493-3

Review article

Check for updates

# Bridging individual-level and system-level approaches to advance psychology-based diversity initiatives

Joyce C. He [1], Benjamin B. Keller [1] & Sonia K. Kang [2]

## Abstract

For decades, researchers, leaders and policymakers have worked to develop and implement interventions to increase the organizational representation of historically under-represented and marginalized groups, such as women in STEM and Black students attending prestigious universities. Despite substantial investments of time and resources, progress has stalled — and, worryingly, these efforts are facing growing backlash. In this Review, we examine diversity initiatives and policies grounded in psychological theory, particularly social cognition and person perception. We begin by outlining common organizational diversity strategies, identifying their psychological foundations and assessing their effectiveness. Although these approaches address an essential dimension of under-representation, they have limited effectiveness when applied alone because they primarily target individuals and intrapersonal processes (for example, stereotypes, prejudice and discrimination) while leaving systems that perpetuate inequality intact. We then consider adjacent literatures of choice architecture and judgement and decision-making, which offer complementary tools for advancing diversity by addressing both the systems in which people operate and the processes that shape individual behaviour. When combined with psychologically informed initiatives, these approaches offer a promising and sustainable path towards meaningful progress in organizational diversity.

## Sections



[1]Anderson School of Management, University of California, Los Angeles, Los Angeles, CA, USA. [2]Department of Management, University of Toronto Mississauga, Mississauga, Ontario, Canada. e-mail: sonia.kang@utoronto.ca

## Introduction

Decades of research across diverse fields including sociology, economics and psychology have investigated inequality in organizations and labour markets from different theoretical perspectives. One perspective that has been particularly influential is social cognition (a subfield of social psychology), which examines how people process, store and apply information about others and social situations to form perceptions and judgements, which ultimately influence behaviours and decision-making in social contexts[1,2]. A key theoretical tenet of social cognition is that such social-cognitive processes are fundamentally framed by social categories, such as race, gender and age[3–5]. Categorization is typically automatic and influences perceptions, attitudes and behaviours. Thus, social-cognitive theories provide a foundational framework for explaining the lack of diversity in organizations by highlighting how workplace decisions and behaviours — especially those made during key entry points (such as employee attraction, recruitment and selection) and pathways to advancement (such as training and development, promotions, mentorship and sponsorship) — are shaped by category-based perceptions and cognitions. Importantly, these insights also inform diversity initiatives, that is, policies, interventions and practices aimed at increasing organizational diversity (the representation of individuals from historically under-represented or marginalized identity groups across all organizational levels).

Despite the widespread adoption of diversity initiatives grounded in these psychological insights, evidence on their effectiveness is mixed[6–8]. These efforts might fall short because psychology-based diversity initiatives focus too heavily on individual-level behaviour change, reflecting psychology's emphasis on intrapersonal processes. This individual-level focus overlooks the powerful influence of social, organizational and systemic factors on behaviour change. By contrast, the field of behavioural design, which applies insights from psychology, behavioural economics and decision science to influence human behaviour, is grounded explicitly in the core tenet that behaviour is shaped by the choice environment, and that changing the choice environment can influence behaviour without changing the psychology of the individual or the content of the choice. Thus, diversity initiatives rooted in behavioural design might better reflect how behaviour is shaped in organizations than individual-focused solutions alone and might therefore be more effective. Underscoring that potential, a large-scale field experiment found that behaviourally designed diversity training increased hiring of women and non-national applicants[9].

In this Review, we present behavioural design as a promising avenue for bridging individual-level and system-level approaches to organizational diversity. We begin by outlining common psychology-based diversity initiatives and evaluating their effectiveness in improving organizational diversity outcomes. Next, we critically assess this body of work and highlight its predominant — and often limiting — focus on individual and intrapersonal processes. To address this limitation, we propose a behavioural design approach that integrates the systems and contexts in which people operate and the individual psychological processes that influence behaviour. We review the literature applying behavioural design to diversity initiatives and their limitations. We conclude by summarizing key insights and providing a roadmap for future research and practice.

## Psychology-based diversity initiatives

In this section we provide an overview of initiatives that leverage foundational psychological theories in social cognition to increase diversity in organizations. These approaches engage either perceivers or targets by reducing biased perceptions and decisions among evaluators or by fostering the entry and participation of members of under-represented groups, respectively. Note that some theories that support perceiver-focused initiatives might also apply to target-focused initiatives, and vice versa, and many social-cognitive theories overlap, even across the groupings we present. Although multiple psychological theories are likely to inform any one diversity initiative, for clarity we organize them according to the theory that most centrally informs their design.

### Perceiver-focused initiatives

Perceiver-focused initiatives (Table 1) often involve diversity training[10], which varies widely in its format and content. Below, we summarize key groups of psychological theories and the specific practices they inform, which serve as foundational elements of diversity training.

**Understanding bias.** Social categorization processes enable humans to use mental shortcuts and heuristics that streamline cognitive processing in a complex social world[11,12]. The natural tendency to categorize social information leads to the development of generalized beliefs (stereotypes), attitudes (prejudice) and behaviours (discrimination). Although prejudice typically refers to negatively valanced attitudes, the content of stereotypes can be more nuanced and mixed. According to the stereotype content model[13–16], stereotypes have two primary dimensions (warmth and competence) that characterize generalized beliefs about groups and influence downstream attitudes and behaviours.

Importantly, the concept of implicit social cognition[17–20] suggests that social-cognitive processes usually operate automatically, outside conscious awareness. Research based on large-scale data collection through online implicit association tests demonstrates the persistence of implicit attitudes and stereotypes, reflected in automatic associations[21]. For example, there are automatic associations between Black people and negative evaluations, and White people and positive evaluations, as well as automatic associations between men and science, and women and the humanities.

The dual-process model of prejudice[17,22–24] distinguishes between implicit, automatic cognitions and more explicit, controlled beliefs. Importantly, according to this model individuals can consciously counteract automatic processes by engaging in more deliberative and controlled thinking, enabling them to exert more influence over their behaviour[25,26].

Several diversity initiatives are informed by these theories. Implicit bias training educates employees about the nature and origins of implicit bias, with the goal of raising awareness and making this automatic process controllable. Training is usually facilitated by an expert who teaches employees to recognize their own biases, often by completing and receiving feedback on an implicit association test.

By contrast, stereotype awareness training targets explicit beliefs and stereotypes. The goal is again to raise awareness, with the hope that recognizing biased beliefs will prompt individuals to consciously monitor and suppress stereotypical thinking, thereby reducing its influence on behaviour.

In addition to raising awareness, some forms of diversity training use various strategies to actively reshape biased beliefs. Common approaches involve attempts to weaken stereotypes by exposing individuals to counter-stereotypical exemplars[27–29] or using contradictory pairings to weaken implicit associations[30,31].

Finally, bias feedback training (commonly used in implicit bias or stereotype awareness training) involves providing trainees with

**Table 1 | Psychology-based diversity initiatives that target perceivers and evaluators**

| Psychological focus | Key theories | Intervention type | Description | Representative evidence |
|---|---|---|---|---|
| Understanding bias | Implicit social cognition[17–19]<br>Dual-process model of prejudice[22,23]<br>Stereotype content model[13]<br>Social categorization[11,12]<br>Exemplar theories of social judgement[27] | Implicit bias awareness training | Educates employees about how biases operate to increase awareness and encourage self-regulation and/or change of biased cognitions (stereotypes), attitudes (prejudice) and behaviours (discrimination) | A meta-analysis of 208 effect sizes found modest reductions in biased beliefs ($g$=–0.20) compared with a pre-intervention baseline, with negligible behaviour change ($g$=0.08)[108] |
| | | Stereotype awareness training | Raises awareness of stereotypes to encourage conscious monitoring and suppression, thereby reducing discrimination | A large field experiment ($n$=3,016) in an international organization found that diversity training (including stereotype awareness and feedback) improved attitudes towards women and increased both awareness of gender bias and gender-inclusive behavioural intentions ($d$=0.13–0.21) compared with a control group who received training unrelated to diversity ($d$=0.01–0.03), but did not lead to lasting behaviour change measured 3–14 weeks later ($d$=0.01–0.03[a])[109] |
| | | Implicit bias/ stereotype change | Reshapes implicit associations or stereotypes using various methods including training new associations and engaging with counter-stereotypical exemplars | An internal meta-analysis ($n$=17,000 undergraduates) found that evaluative conditioning ($d$=0.21) and vivid counter-stereotypes ($d$=0.49) briefly reduced implicit bias compared with a control condition that did not receive any intervention, but did not significantly reduce explicit bias ($d$=0.04)[107] or have lasting effects 2–4 days later ($d$=0.08–0.05)[110] |
| | | Bias feedback training | Provides personalized feedback about biases to motivate behaviour change | A meta-analysis found that bias feedback training improved explicit attitudes ($d$=0.45), but had smaller and less consistent effects on behavioural intentions ($d$=0.06), compared with a control group who received training unrelated to diversity[7]; these interventions have not been tested in the field |
| Intergroup relations | Intergroup relations and conflict[36,37]<br>Intergroup contact[11,36,46] | Perspective-taking exercises | Encourage individuals to take the perspective of someone from a different group | Meta-analyses suggest that perspective-taking leads to modest effects on behavioural change ($g$=0.23) and shifts beliefs, attitudes or emotions ($g$=0.18) compared with a pre-intervention baseline[108] |
| | | Relationship-building exercises | Facilitate intergroup contact between groups by inviting employees from various backgrounds to engage in dialogue | A meta-analysis suggests that imagining an interaction with someone of a different race has a small effect on explicit attitudes ($d$=0.12) across studies (mostly with college students)[7]<br>Dialogue interventions modestly reduced prejudice ($d$=0.2) compared with other interventions, and showed stronger behavioural effects 1 day after ($d$=0.31) compared with immediately after ($d$=0.26) the intervention[7] |
| | | Cultural competency training | Equips employees with skills for effective and harmonious cross-group interactions | Online cultural competence training increased cultural self-efficacy and knowledge ($d$=0.12) compared with a control condition that did not complete the training; behaviour was not measured[129] |
| Self-regulation and goal pursuit | Self-efficacy[60]<br>Theories of planned behaviour[58]<br>Consistency theories[33,34] | Bystander intervention training | Teaches employees how to intervene when they witness bias, discrimination or harassment | Workplace bystander intervention training led to small gains in bystander efficacy ($d$=0.09) and intent to intervene ($d$=0.09) measured 2 weeks after the intervention compared with a pre-intervention baseline, but no lasting changes in attitudes 1 month later ($d$=0.31–0.47, results are non-significant owing to high attrition and the small sample size)[130] |
| | | Implementation intentions | Encourages employees to create 'if–then' plans that link situational cues ('if') with goal-directed behaviours ('then') | Implementation intentions reduced implicit racial bias ($d$=0.38)[107] compared with a control condition that did not receive any intervention but did not decrease explicit bias ($d$=0.04) or show lasting effects 2–4 days later ($d$=0.08)[110] |

Effect sizes are presented if reported in the original paper or calculable from the published material. $g$, Hedges $g$ value; $d$, Cohen's $d$ value. [a]The effect size was calculated by the authors of the current Review.

personalized feedback about their biases to encourage and motivate personal change[32–34]. When paired with training that raises awareness of bias and strengthens intentions to reduce discrimination, this approach can reduce stereotype activation and application[35].

**Intergroup relations.** Intergroup theories examine how individuals from different social groups perceive, experience and interact with one another, highlighting how intergroup differences shape dynamics such as conflict and cooperation[36,37]. These dynamics are driven by social-cognitive processes (including implicit cognition and dual-process models of prejudice) that shape how people think, feel and behave towards in-groups versus out-groups[37–39]. Social identity theory and self-categorization theory explain intergroup conflict as stemming from an 'us versus them' mindset[40].

Diversity initiatives rooted in theories of intergroup relations include perspective-taking exercises, relationship-building and dialogue exercises, and cultural competency training. Perspective-taking exercises encourage participants to actively consider the mental

states and subjective experiences of individuals from different backgrounds. For instance, individuals might be asked to put themselves in the position of an out-group member by writing a short essay from their perspective[41–44]. Other activities to facilitate perspective-taking and foster empathy include role-playing, simulations and reflective exercises.

Intergroup contact theory suggests that direct contact between members of different groups can reduce prejudice and improve intergroup relations[36,45–49]. Some diversity training applies this concept by facilitating direct contact between groups, inviting employees from diverse backgrounds to engage in dialogue or collaborate on tasks that require teamwork to achieve shared goals[50,51]. These exercises aim to promote meaningful conversations, build connections and foster cooperation.

Finally, cultural competency training is designed to provide employees with the skills needed for effective and harmonious cross-group interactions. Rather than changing attitudes or increasing motivation to engage in such interactions, this training focuses on providing tools for effective communication and collaboration across diverse cultures and backgrounds, such as active listening and empathy.

**Self-regulation and goal pursuit.** Another influential group of theories focuses on intrapersonal processes involved in setting, pursuing and achieving goals for social interactions, with broad applications to behaviour change and self-regulation in social settings. Goal setting theory[52–55] suggests that specific, challenging goals increase motivation and goal attainment. Building on these insights, the theory of planned behaviour[56] proposes that behaviour change is more likely when people set concrete, specific goals and plan for the contexts in which they will activate and engage in these behaviours[57–59].

Success in planned behaviour is further supported by self-efficacy[60–63] (the belief in one's ability to succeed) and by ensuring that the behaviours align with an individual's beliefs, values and previous actions[33,64]. Together, these theories suggest that individuals with clear intentions and positive attitudes towards diversity can use self-regulation and goal setting to reduce biased decisions and behaviours.

Building on these seminal theories, bystander intervention workshops and training aim to teach employees how to intervene in situations where they witness bias, harassment or discrimination. This training often focuses on reducing psychological barriers to intervention and action, and increases self-efficacy to empower observers to become active allies.

Implementation intention initiatives build on the theory of planned behaviour by encouraging employees to create 'if–then' plans that link situational cues ('if') with goal-directed behaviours ('then') to increase the likelihood of allyship (for example, 'if I notice someone being ignored in a meeting, I will help amplify their ideas by directing attention back to them') and reduce undesirable behaviour (for example, 'if I hear someone make an inappropriate joke, I will address it by calmly stating that such comments are unacceptable').

## Target-focused initiatives

Target-focused initiatives (Table 2) are designed to support current or prospective employees from under-represented groups and to increase their representation, particularly in roles, occupations or fields where they remain marginalized. Here we summarize key groups of psychological theories that are commonly used to explain under-representation (some of which also inform perceiver-side initiatives), along with the diversity initiatives they inform.

**Identity threat.** Social identities refer to the social categories and groups that people belong to and how those identities shape beliefs, attitudes and behaviours[37,65–67]. Employees from under-represented or marginalized groups might experience social identity threat — a concern or worry that their group identity might be devalued or stigmatized within a given context. This threat can lead to avoidance, withdrawal and intentions to exit particular fields, organizations or roles[68–70].

**Table 2 | Psychology-based diversity initiatives that target individuals from under-represented and/or marginalized groups**

| Psychological focus | Key theories | Intervention type | Description | Representative evidence |
|---|---|---|---|---|
| Identity threat | Social identity theories[37]<br>Identity/role conflict[81,83]<br>Stereotype threat and social identity threat[76]<br>Self-efficacy and confidence[60] | 'Lean in' training | Empowers individuals from under-represented groups to build confidence, improve skills and practice counter-stereotypical behaviours | An internal meta-analysis across six experiments showed that 'lean in' messages increased perceptions of women's empowerment compared with a control condition, but also strengthened views that women have caused and should fix gender inequality[146] |
| | | Self-affirmation intervention | Uses brief self-affirmation exercise to boost self-esteem and reduce social identity threat | An internal meta-analysis of two field studies showed that self-affirmation reduced the gender gap in MBA course performance ($g$=0.14) compared with a control condition ($g$=0.54); this effect was mediated by lower self-doubt among women in the self-affirmation condition[140] |
| Social influence and norms | Social learning theory[93]<br>Social comparison theories[95]<br>Self-categorization theory[38]<br>Social norms[96] | Mentorship programmes | Pairs individuals from under-represented groups with mentors or sponsors and encourages regular meetings | Female undergraduate engineering students with female mentors maintained a more stable sense of belonging in engineering than those with no mentors or male mentors; both of the latter groups showed decreases in belonging compared with students with female mentors ($d$=0.33[a])[147] |
| | | Employee resource and affinity groups | Social groups for specific demographic or identity groups | Membership in LGBTQ+ employee resource groups was associated with greater identity disclosure at work ($r$=0.12), but only marginally associated with lower perceived discrimination ($r$=–0.08)[149] |

Effect sizes are presented if reported in the original paper or calculable from the published material. $g$, Hedge's $g$ value; $r$, Pearson's $r$ coefficient; $d$, Cohen's $d$ value. [a]The effect size was calculated by the authors of the current Review.

Social identity threats often manifest as a lack of social fit and belonging[69,71–74] and might also include concerns about performance and competence, especially when individuals fear confirming negative stereotypes about their group (stereotype threat)[68,75,76]. These threats can spill over into other workplace experiences and behaviours[77], reducing self-efficacy and confidence[60,78]. Social identity threat can be deeply entrenched, as individuals internalize domain-specific self-stereotypes from a very young age[79,80].

Individuals hold multiple social identities, and the interactions between these identities can create conflict or tension[81,82]. For instance, women might see themselves as ill-suited for leadership roles because norms associated with their gender identity conflict with those linked to the traditionally masculine prototype of a leader[83]. This misalignment underpins several models of inequality that focus on identity or role incongruence and conflict[80,83–85].

One common type of diversity training that is based on these concepts aims to empower individuals from under-represented groups by equipping them with psychological resources, skills and a sense of agency. The goal is to boost confidence and encourage engagement in career-advancing behaviours such as negotiating, self-advocacy and applying for promotions. Often referred to as the 'lean in' approach[86,87], this type of training has become widespread, particularly for women. Its popularity is reflected in the proliferation of women's leadership and career development workshops, which seek to address systemic barriers by fostering individual empowerment and self-efficacy.

Finally, self-affirmation interventions draw directly from self-affirmation theory[88–92] and involve brief exercises such as written reflections that encourage individuals to affirm core values or personal strengths. These exercises aim to boost self-esteem and mitigate the harmful effects of social identity threats. By reinforcing a sense of self-worth, self-affirmation interventions help individuals to navigate identity-related challenges and make decisions that align with their goals, rather than avoiding behaviours or opportunities because of perceived threats or stigma.

**Social influence and norms.** The final group of theories emphasizes how people use social cues to guide their behaviour and shape self-evaluations. According to social learning theory[93], people learn behaviours by observing and imitating others. Building on this concept, social comparison and social-observational theories[61] propose that people look to role models for guidance, aspiration and self-evaluation[94,95]. More generally, people learn the 'unwritten rules', or norms, of appropriate behaviour by observing others, especially members of their own social groups (that is, their in-groups)[96–100]. Social norms can also be determined by the broader group context, such as workplace culture (defined as the system of shared values and norms that define appropriate attitudes and behaviours for organizational members[101]), which can also influence which attributes and behaviours are rewarded or penalized in an organizational setting (Box 1). Together, these theories highlight how social factors such as the absence of mentors or referent groups who share one's identity can limit access to norms, aspirations and behavioural models, thereby contributing to the under-representation of certain groups in organizations[80,102].

Diversity initiatives based on social influence and norms include mentorship programmes and employee resource or affinity groups. Formal and informal mentorship programmes are designed to support the career development of individuals from under-represented groups by connecting them with mentors for ongoing guidance and support. Mentors help mentees navigate workplace challenges, set goals, and build confidence and competence. When mentors share the mentee's identity, they can also serve as powerful role models, reinforcing a sense of belonging and improving retention. By fostering these relationships, mentorship programmes help address both career-related and identity-related barriers, contributing to a more inclusive and supportive organizational environment.

Employee resource groups and affinity groups are designed to build community and foster a sense of belonging among employees who share a common identity (for example, women, people of colour or LGBTQ+ people). These groups aim to increase retention and representation across organizational levels by creating supportive spaces where members can share experiences, seek advice and build professional networks. Seeing peers with similar identities in the organization boosts employees' career aspirations, sense of belonging, competence, and commitment to stay and advance careers.

## Effectiveness of psychology-based diversity initiatives

Having reviewed key psychological theories and their corresponding diversity initiatives, we now provide a high-level summary of their effectiveness across key indicators of organizational diversity (for example, retention, representation and advancement of under-represented or marginalized groups). Representative evidence for specific diversity initiatives is included in Tables 1 and 2.

### Effectiveness of perceiver-focused initiatives

Most research on perceiver-focused initiatives has focused on diversity training, broadly defined as a range of interventions that commonly involve implicit bias awareness and stereotype awareness training. This research has yielded mixed results[103,104]. Meta-analyses[10,105], large-scale reviews[106] and multi-site laboratory experiments[107] suggest that diversity training can successfully foster short-term improvements in attitudes, motivation and skill retention. However, evidence from meta-analyses[105,108], a field experiment with real employees[109] and laboratory experiments[110] suggests that these effects rarely lead to the sustained behavioural change required to meaningfully improve organizational diversity. Notably, reductions in implicit bias observed after training often rebound within days or even hours among undergraduates in a laboratory setting[110]. Longitudinal survey research of employees across US organizations similarly shows little evidence that diversity training increases the representation of employees from minority groups in managerial roles[6].

When positive effects of diversity training do occur on attitudes (and, more rarely, behaviours), they are typically context-dependent, such that they vary by target group[106] or based on participants' pre-existing attitudes and beliefs[109,111]. In addition, meta-analytic evidence from studies in the laboratory and field suggests that diversity training tends to be more effective when it is intensive and tailored to the organizational context. Specifically, diversity training is more effective at changing attitudes and behaviours when it is delivered over an extended period, includes distributed practice (sessions spaced out over time), is conducted in person rather than online, combines awareness and skills-based training, provides opportunities for social interaction and is supported by complementary diversity initiatives (such as organizational messages that signal explicit support for diversity)[10,105].

Moreover, diversity training initiatives can backfire, triggering reactance, decreasing diversity or leading to other unintended consequences[112–114]. For example, in laboratory research with students and online participants, diversity training reduced sensitivity

## Box 1 | Workplace culture

Organizational culture refers to the system of shared values and norms that define appropriate attitudes and behaviours for members of an organization[157,222–224]. Organizational culture not only influences how individuals behave and interpret interactions but also reflects shared norms and practices that can reinforce inequality. Thus, organizational culture plays an important role in shaping organizational gender diversity[157,223–225].

### Masculine cultures and signals of belonging

The specific content of organizational culture — which values and behaviours are treated as normal and rewarded — shapes who feels as though they belong and can succeed. In many US organizations, stereotypically male traits such as assertiveness and dominance are treated as the standard for success[71,72,224]. These 'masculine defaults' are often embedded in policies, internalized by employees and reinforced over time. They send a powerful signal about who is welcome and likely to succeed, which might deter women from entering or advancing in certain fields[71,72,225]. Scholars have described the workplace as a masculinity contest where employees compete for power by displaying masculine traits[223,226]. These dynamics can also emerge indirectly; for example, given societal associations between brilliance and masculinity, workplace cultures that prize brilliance might foster masculinity contests and undermine women's sense of belonging[227,228].

### Culture as a moderator of diversity, equity and inclusion efforts

Organizational culture also shapes how employees respond to diversity efforts. Field research shows that these efforts are most successful when they align with the organization's broader cultural values[105,229]. By contrast, when diversity initiatives are perceived as conflicting with the dominant culture, they might be seen as a threat, particularly by advantaged individuals who benefit from the status quo[230]. Employees in inclusive organizational climates are less likely to resist diversity training because they see it as aligned with the organization's core goals rather than as an external imposition[106,231].

### Culture as a target of intervention

These insights suggest new directions for targeting workplace culture as a lever for increasing diversity. Cues signalling inclusion (such as inclusive policies or high representation of women) can reduce social identity threat and increase belonging among women compared with cues signalling lack of inclusion through low representation of women[232]. Similarly, removing cues that signal a masculine culture (such as gendered language in job adverts or emphasis on brilliance over growth) can increase women's interest in and persistence within male-dominated fields[193]. Although promising, this line of work is relatively nascent. More empirical research is needed to understand what drives cultural change and to test interventions that can shift organizational culture over time. Because culture change unfolds gradually, longitudinal research embedded in real-world organizations is essential[233].

### Implications for psychology and behavioural design

A cultural lens enhances both psychology-based and behavioural design approaches by highlighting how shared norms, values and institutional practices shape individual cognition and behaviour. This perspective helps explain why interventions might succeed in one setting but fail in another. For instance, efforts to reduce bias or promote inclusion might be ineffective if they conflict with dominant cultural norms. Behavioural design offers a practical entry point for initiating cultural shifts by altering interaction patterns or default practices; in turn, these behavioural changes can help to reshape the broader culture[193,234]. These perspectives are complementary: culture shapes the psychological processes targeted by behavioural interventions, and behavioural design offers tools to begin shifting cultures from the ground up. Recognizing this synergy is key for understanding when and how cultures can truly change to become more inclusive.

to discrimination among individuals in the dominant group (compared with individuals who completed separate training unrelated to diversity) by reinforcing the belief that the systems to prevent discrimination are already in place[115,116]. Evidence from laboratory studies of MBA students and fully employed individuals shows that bias awareness training can also inadvertently exacerbate stereotype expression and discriminatory behaviour by unintentionally signalling that biases and stereotypes are normative and therefore permissible[117]. Finally, evidence from laboratory studies of undergraduates[118] and field surveys of employees across US organizations[6,118,119] suggests that mandatory diversity training or training perceived as unfair can provoke resistance or backlash among dominant or majority group members. These unintended consequences help explain the overall mixed evidence on the effectiveness of diversity training: although these initiatives might positively shift attitudes, their positive effects might be offset by the fact that benefits are limited to certain groups[110] or accompanied by negative outcomes such as backlash[112], increased prejudice[117] or decreased diversity[119].

Empirical research examining the effectiveness of other perceiver-side initiatives shows a similar pattern: modest improvements in explicit attitudes[7] but limited and inconsistent effects on behaviour[7]. When positive outcomes on explicit attitudes, momentary behaviour or behavioural intentions are observed, they are typically context-dependent, and studies rarely assess actual lasting behavioural change. In particular, light-touch interventions (such as perspective-taking or imagined contact) have grown in popularity over the past 20 years[7], despite leading to only modest improvements in explicit attitudes and behavioural intentions[7,120,121] with little impact on implicit attitudes[110], according to robust meta-analytic evidence from the laboratory and the field. The effects of such light-touch, intergroup-based interventions on actual behaviours remain unclear, given the limited research that measures behavioural outcomes in the field directly[7,122]. A large-scale meta-analysis suggests that more involved forms of intergroup contact (such as dialogue and in-person interaction) show more promise, yielding modest but long-lasting behaviour change[7]. However, other meta-analyses suggest that intergroup contact can also have unintended consequences, such as suppressing majority group members' recognition of discrimination and reducing their intentions to engage in collective action[47,123].

Finally, implementation intention interventions reduce implicit bias compared with a control condition in laboratory settings[107,124,125], but these interventions have not been widely tested in the workplace. The limited research that exists has not found them to be effective in decreasing gender bias in promotion decisions[126], although they have been shown to be effective for more general behavioural and habit change in other domains[127,128]. There is limited research evaluating bystander training and cultural competency training in workplace settings. The evidence so far from both laboratory experiments[129] and field surveys[130] suggests that these interventions can produce short-term improvements in attitudes, skill-related efficacy and intentions, but these effects fade over time[7,129,130]. Few studies have examined actual behavioural outcomes[131]; although some early research suggests that bystander training might influence anti-discriminatory behaviour among faculty in academia, this evidence is limited by small sample sizes and a reliance on self-reported measures[132]. Additionally, some evidence indicates that bystander training can reduce sexual harassment compared with receiving no bystander training, particularly in college and medical settings[133,134].

### Effectiveness of target-focused initiatives

Evidence on the effectiveness of target-side initiatives is also mixed[135]. Although such initiatives are quite common (especially 'lean in'-style training), rigorous empirical evaluations remain limited. Some research (primarily conducted using student samples in the laboratory) suggests that women's confidence might improve when they adopt a gender-neutral outlook[136], develop a growth mindset towards agentic behaviours and roles[137,138], reject negative stereotypes and engage in self-affirmation[139,140]. Although these strategies have been linked to increased empowerment and self-efficacy, little research has examined their effects on objective outcomes such as increased representation and retention. Among the few studies that do examine behavioural outcomes, a self-affirmation intervention attenuated the gender gap in performance between male and female MBA students[140]. Affirmation-based interventions might also facilitate social norm change, which might foster equality over time[141].

Furthermore, encouraging women to adopt self-promotional and career-advancing behaviours might expose them to unintended career penalties[87]. Seminal research on prescriptive stereotypes shows that women and individuals from racial and/or ethnic minority groups often face backlash for defying identity-based expectations. For example, female leaders are often penalized for being assertive, yet criticized when they are not[142–144]. Women who downplay femininity to signal fit in male-dominated workplaces are paradoxically less likely to be hired because they violate norms according to which women should maintain femininity[145]. Furthermore, 'lean in' messaging can unintentionally shift the responsibility for both creating and mitigating gender inequality onto women themselves[146], reinforcing the idea that individual effort — rather than systemic or organizational change — is the key to achieving equity.

Mentorship programmes and employee resource groups have been studied more than other target-focused initiatives and show more promise, but their effectiveness is context-dependent. According to meta-analytic evidence from the laboratory and the field, mentorship and networking initiatives are more consistently beneficial for boosting target employees' positive experiences and satisfaction than for increasing representation or promotion rates[6,106]. Outside organizational settings, field experiments among college students suggest that peer mentorship can improve retention and experiences among under-represented groups[147], but such initiatives have not been tested in workplace settings. Survey research on employees across US organizations has linked employee resource groups and affinity groups to greater perceived inclusion and comfort with disclosing minority identities[148,149]; however, evidence of their impact on behavioural outcomes or representation is limited and inconsistent[6], and large-scale quantitative evaluations of their effectiveness are scarce[150].

## Why psychology-based diversity initiatives fall short

Our Review reveals that most evaluations of psychology-based diversity interventions prioritize short-term cognitive indicators such as shifts in attitudes[7] rather than observable behavioural outcomes. Although attitude change is an important metric, it does not necessarily translate into the sustained behavioural shifts required to improve diversity outcomes. Echoing previous reviews[7,103], we find that only a small proportion of studies assess behaviour change, and even fewer examine long-term behavioural impacts. This lack of focus on long-term behavioural impacts is important, because interventions that show promise for shifting attitudes often fail to generate meaningful, lasting behaviour change[109,110,151] — although attitude change might lead to behaviour change, it is not a guaranteed outcome.

Additionally, much of the evidence supporting these initiatives (including studies validating their effectiveness) comes from laboratory or online settings using convenience samples. These controlled settings differ substantially from the organizational environments where these initiatives are intended to be applied. A common but flawed assumption is that behaviour change stems from stable intrapersonal processes that should therefore generalize across settings. However, the disconnect between experimental settings and real-world application is likely to contribute to the limited effectiveness observed in practice. Indeed, the existing research reviewed above suggests that the benefits of training often do not translate into actual workplace behaviour.

A broader challenge with psychology-based approaches lies in their heavy reliance on individual-level solutions to drive behaviour change. Although these interventions play an important role in inducing change, focusing exclusively on intrapersonal processes oversimplifies the complex nature of behaviour and overlooks the powerful influence of social environments, organizational contexts and systems.

Perceiver-focused interventions are designed to change behaviour by reshaping how people think, such as weakening stereotypes or increasing awareness of biased information processing. The underlying assumption is that awareness will motivate people to control their biases and modify their behaviour. However, core findings of social cognition research demonstrate that psychological phenomena such as implicit bias, stereotypes and identity-related perceptions are automatic and deeply ingrained, and stem from basic social perceptual processes that are 'hard-wired'[12]. Although individuals can try to suppress or override these implicit associations[22,23,124], approaches that rely solely on intrapersonal efforts are often insufficient[152], especially when they do not also account for the broader social and environmental forces that shape behaviour[153–155].

Target-side interventions similarly often focus on the individual by, for instance, encouraging women and members of other marginalized groups to 'lean in'[86,87,146], develop their competence and confidence, and adopt behaviours associated with dominant groups (such as men, White people or higher-status individuals)[156–158]. These interventions again have an intrapersonal focus: they aim to motivate targets to

alter their psychology, overcome internal doubts and change their behaviour. However, these approaches overlook how identity-related behaviours and perceptions are shaped through lifelong socialization and constrained by systemic factors[79,159,160], such as prescriptive norms that dictate how certain groups are expected to behave and the penalties for deviating from these expectations[143,144,161]. Moreover, even when individuals do change their behaviour, systemic and environmental barriers often remain in place and do not diminish in strength.

Our Review also underscores that diversity initiatives that foster meaningful social interaction and deep engagement (such as cross-group interactions, dialogue-based relationship-building exercises and in-person diversity training that emphasizes interaction) tend to produce more consistent and long-lasting changes in both attitudes and behaviours than lighter-touch interventions (such as imagined contact or online diversity training modules)[7,10,162]. However, these more involved initiatives remain relatively rare in organizational settings[7], probably because they demand more time, effort and resources — factors that might deter organizations from prioritizing them despite their demonstrated impact.

Importantly, the reasons that psychology-based diversity initiatives are ineffective are not inherent flaws of psychology as a discipline nor flaws of its foundational theories. The emphasis on attitudes, cognition and intrapersonal processes — and therefore individual-level solutions — reflects psychology's orientation as a basic science that proposes and tests fundamental mechanisms under controlled conditions with the aim of providing a generalizable explanation for phenomena related to the human mind; psychological theories are not typically developed with the direct goal of producing applied solutions. That is not to say that psychological theory cannot and does not encompass intervention-based research; rather, the majority of the field — including the foundational theories based in social cognition that we have explored here — focus on explaining phenomena rather than informing policy.

The issue, then, lies in the challenges of translating and applying basic psychological theories directly to organizational policy and practice. Such translation from theory to policy often skips important intermediate steps, such as developing applied models based on basic theory, testing behavioural outcomes in the laboratory and adapting interventions to organizational contexts. This issue of translation is often compounded when those applying the theory differ from those who developed it, increasing the risk of poor fidelity and low intervention effectiveness.

Nonetheless, even if shortcomings of diversity initiatives arise from the complexities of translation rather than flaws in the theory or discipline itself, they do not undermine the core limitations we identify. The disproportionate reliance on individual-level approaches to diversity has important implications: focusing on individual change shifts the burden of progress onto individuals — especially those from marginalized groups — rather than organizations[163]. Expecting marginalized employees to change their behaviour without corresponding systemic efforts to remove structural barriers reinforces inequality rather than eliminating it. To achieve meaningful, lasting diversity, organizations must move beyond individual-level interventions and address the systemic and contextual factors that perpetuate inequity.

## A behavioural design approach to diversity

Although our critique highlights the limitations of individual-level approaches, we are not suggesting that they should be abandoned altogether. Rather, their limited effectiveness stems from their isolation, that is, when individual-level solutions are implemented without also addressing the systems and contexts that shape behaviour. Thus, diversity initiatives should move beyond the current focus on individuals and their psychological processes and integrate perspectives from related disciplines — such as sociology and economics — that illuminate system-level influences on behaviour.

Sociology and economics emphasize macro-level systems rather than intrapersonal processes. Solutions to increase organizational diversity informed by these approaches (such as affirmative action or equal employment mandates) typically involve heavy-handed and substantive policy-level changes[6,106,164]. These strategies aim to change incentives or bypass individual choice and agency altogether to produce large-scale shifts in behaviour and representation. Although such strategies are impactful, they come with their own limitations and criticisms. Specifically, they often overlook individual agency and autonomy, and might generate resistance or non-compliance[119].

Thus, the most effective diversity initiatives must consider both levels: the systems in which people operate and the individual psychological processes that influence behaviour. One promising framework that integrates these perspectives and acknowledges their interconnectedness is behavioural design — a growing field that draws from psychology, behavioural economics and decision science to understand how the design of choice environments shapes individual decision-making and behaviour.

Rather than focusing on what decisions people make, behavioural design examines how those decisions are presented, that is, how options are framed, how information is conveyed and how incentives are structured. A central concept in this field is that a small, strategic change to the choice environment (a 'nudge') can produce substantial behavioural shifts[165]. This behaviourally informed approach has gained widespread attention among academics, practitioners and policymakers, informing policies in health[166,167], finance[168,169] and sustainability[170,171].

Behavioural design holds substantial promise for advancing diversity goals because it brings a systems-aware lens to individual decision-making by showing how the context of choices — including defaults, framing and choice architecture — can be used to reduce cognitive biases and inhibit decisions that create and perpetuate inequality. Specifically, this individual × system perspective helps to generate psychologically informed organizational solutions that redesign decision-making environments and shift underlying systems to facilitate individual behaviour change (Fig. 1). These strengths reflect behavioural design's interdisciplinary roots, as well as its applied focus on influencing behaviour through scientific insight and evidence. By integrating these insights with psychology-based approaches, organizations can move beyond a fragmented strategy to a more comprehensive and effective diversity framework.

For example, consider an organization aiming to increase gender diversity among its senior leadership team (Fig. 2). A psychology-based approach might emphasize intrapersonal processes, such as evaluators' biases against women leaders (perceiver side) or women's lower likelihood of pursuing promotions due to greater risk awareness, uncertainty about leadership ability, and less overconfidence compared with men (target side). These insights often lead to individual-level solutions such as diversity training for evaluators or 'lean in' training for women. By contrast, a macro-level perspective might focus on policy mandates grounded in economic and sociological theories, such as quotas or affirmative action[172,173].

A behavioural design approach offers a middle path that integrates individual-level and system-level perspectives to generate

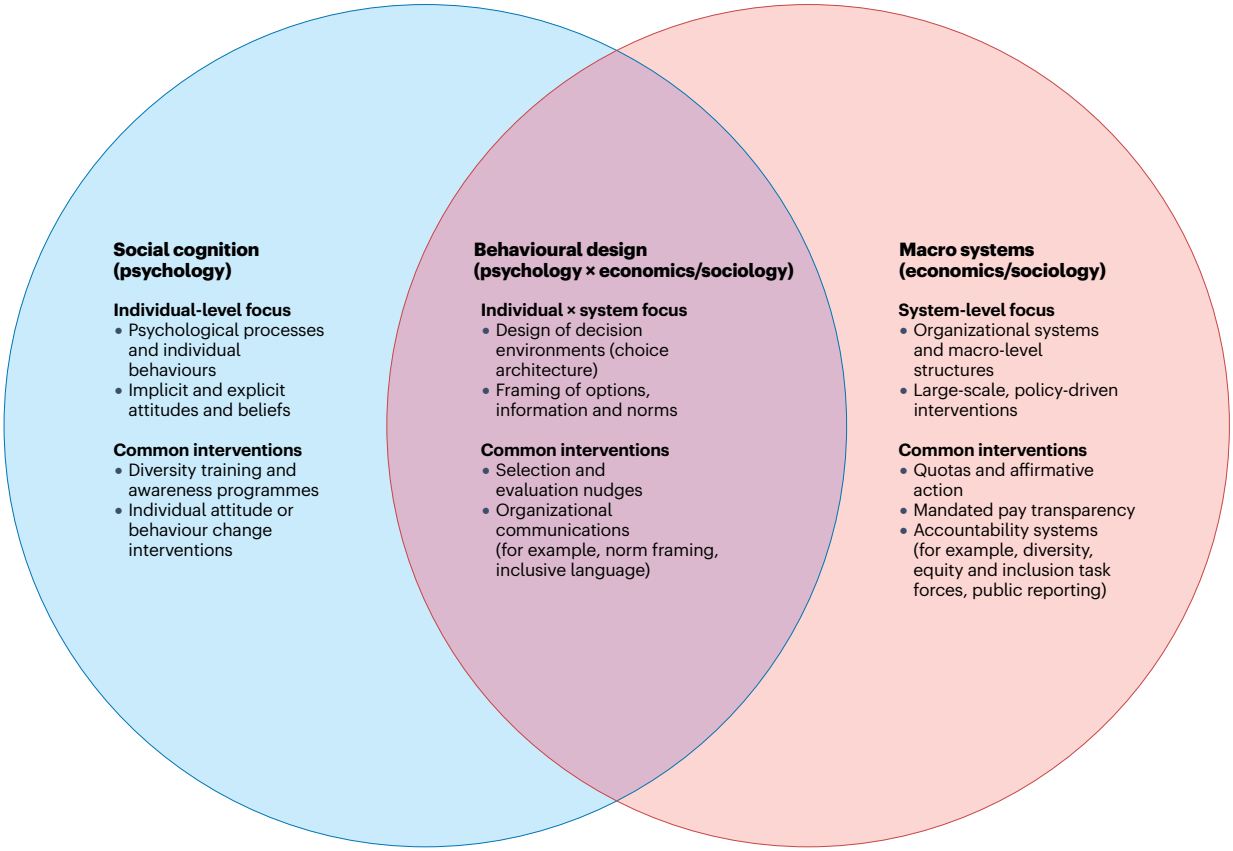


**Fig. 1 | A behavioural design approach to organizational diversity.** Psychology-based approaches to organizational diversity focus on individual attitudes and behaviours (for example, bias awareness training), and macro-level approaches emphasize large-scale structural change (for example, quota mandates). Behavioural design targets the interaction between individuals and systems, and reshapes the decision space through interventions such as choice architecture, norm framing and structural nudges that offer scalable, context-sensitive solutions that preserve individual agency while addressing systemic inequities.

organization-level solutions that are grounded in psychological insights into how individuals perceive and react to choice architecture. For instance, it might identify that the promotion process requires employees to opt in by self-nominating – a decision framing that signals weak or ambiguous norms of participation[174] – exacerbating women's identity-based concerns and doubts[144,175–177]. Reframing the promotion process as opt out – where qualified employees are automatically considered unless they choose not to be – can signal a stronger norm of participation by presenting participation as the default and automatic choice[178,179].

## Key concepts and evidence

A growing body of research supports the effectiveness of such behaviourally informed changes in advancing diversity goals, demonstrating how subtle alterations to the choice environment can leverage inherent psychological processes to guide more inclusive outcomes. For instance, defaults capitalize on the human tendency towards cognitive ease and the powerful influence of perceived endorsement[180,181]. When an option is presented as the default, individuals are more likely to select it owing to low cognitive effort and the implicit signal that it is the recommended or expected choice[178,179]. Indeed, switching promotion processes from opt in to opt out attenuates gender gaps in participation by signalling a stronger norm of inclusion and making participation the path of least resistance[182].

Choice bracketing techniques (such as evaluating candidates jointly rather than sequentially or hiring them in groups) guide evaluators away from individual biases by fostering comparison effects[183]. By presenting multiple options simultaneously, evaluators are encouraged to compare candidates on relevant criteria, rather than making isolated, potentially biased judgements, thereby leading to more equitable assessments and diverse hiring outcomes[184,185]. Even small changes to rating scales (for example, using a six-point scale instead of a ten-point scale) can disrupt implicit associations that might disproportionately favour certain groups, guiding evaluators to more equitable judgements[186]. Similarly, introducing more structure (such as standardizing interview questions or implementing scoring rubrics) can reduce bias by limiting reliance on automatic, intuitive judgements, which are more prone to bias and stereotyping[187]. Instead, adding structure promotes deliberative, criteria-based decision-making, aligning with dual-process models of cognition and prejudice[22].

Similarly, pre-commitment strategies leverage the psychological principle of anchoring, where initial information or decisions heavily influence subsequent judgements[188–190]. By requiring evaluators to define and commit to objective selection criteria before

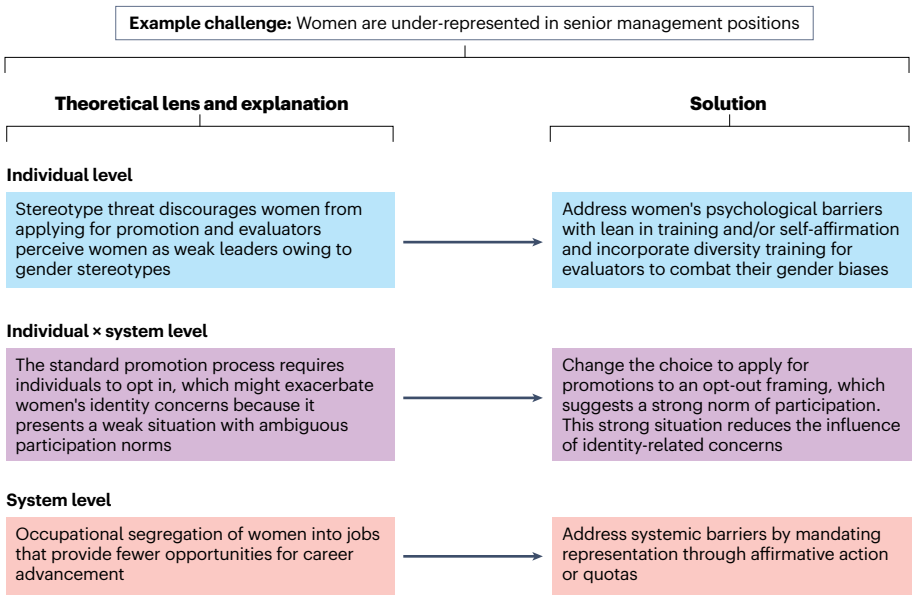


**Fig. 2 | Case study.** Psychology-based (individual-level), sociology-based (system-level) and behavioural design (individual × system-level) solutions to the gender promotion gap whereby women are under-represented in senior management positions.

evaluations, pre-commitment initiatives help to anchor decisions to these pre-established standards, reducing the likelihood of biased post-hoc justifications and thereby promoting fair outcomes[191,192].

Linguistic framing interventions (such as using gender-neutral language in job descriptions) tap into theories of social identity and signalling by removing cues to identity threat, reducing perceived exclusion and boosting a sense of belonging among prospective applicants[193]. Concrete and relatively objective language is less ambiguous, which can increase applications from under-represented groups by making expectations clear and accessible[194].

Information framing can also target implicit bias and stereotype application. For instance, masking identifying demographic characteristics (such as names) on application materials[195] eliminates salient cues that could trigger automatic social categorization processes and activation of implicit biases, leading to more objective evaluations and gender diversity in hiring[196]. That said, the effectiveness of information framing might vary depending on evaluator motivation and organizational context[197,198], and de-identification might limit organizations' ability to track or promote representational equity if used in isolation[199]. Similarly, redirecting evaluators' attention away from potentially biasing information towards job-relevant information can reduce discrimination in decisions on admissions[200].

Social norm nudges explicitly leverage the psychological principles of social influence, including descriptive and injunctive norms[100,201], and can guide behaviour through perceived social expectations. For instance, communicating the pro-diversity attitudes and behaviours of peers or organizational leaders can shift dominant-group members' perceptions and enhance a sense of belonging among individuals from marginalized groups[202]. Thus, individuals might adjust their attitudes and behaviour on the basis of perceived peer norms and leadership expectations[203].

These examples illustrate how behavioural design intervenes at the system level by subtly reshaping the context in which individual psychological processes unfold, steering behaviour towards greater inclusion without requiring direct attitudinal change. Thus, behavioural design offers practical, scalable interventions to address both individual behaviour and decision environments (Table 3). For organizations seeking more effective and sustainable diversity strategies, this approach provides a compelling and evidence-based path forward (see refs. 9,204,205 for reviews that provide additional ideas for interventions based on behavioural design).

## Limitations of behavioural design

Similar to psychology-based approaches, a behavioural design approach to increasing organizational diversity also comes with limitations. First, the empirical evidence on its effectiveness remains relatively limited. Because this is an emerging field, many of the strategies proposed are supported by single studies, in contrast to the meta-analytic evidence available for more established psychological interventions. Thus, more rigorous, systematic research that assesses the effectiveness of using behavioural design to increase diversity across contexts, timeframes and populations is needed. Reviews examining applications of behavioural design in other domains of behavioural change (such as healthcare and financial decision-making) find similar empirical limitations to those we identified in psychology-based approaches: compared with laboratory-based experiments, field-based interventions often produce more modest, context-dependent effects that might fade over time[206–210].

Some scholars also caution that despite its theoretical focus on individual behaviour in context, behavioural design can still lean too heavily on individual-level solutions rather than system-level change, particularly when small nudges and design tweaks leave the broader systems that perpetuate inequality intact[211]. Moreover, although behavioural design aspires to integrate individual psychology with systems-level thinking, in practice some work might focus too heavily on behavioural outcomes at the expense of deeper understanding of the psychological mechanisms driving such change.

In sum, despite empirical limitations, behavioural design is a powerful complement to psychology's individual-level focus by highlighting the systems and environments that shape behaviour. As such, behavioural design should be thoughtfully integrated with psychology-based approaches in existing diversity initiatives.

**Table 3 | Diversity initiatives informed by behavioural design**

| Behavioural design principle | Definition | Diversity initiative | Description | Psychological mechanism | Representative evidence |
|---|---|---|---|---|---|
| Defaults | Using the default choice option to promote a desired outcome | Opt-out framing for competitive processes | Automatically consider qualified candidates for advancement and other opportunities unless they opt out | Leverages cognitive ease, norm signalling and perceived endorsement to guide behaviour[178,217] | In a series of laboratory and field experiments, opt-out framing attenuated the gender gap in participation in competitive selection processes (such as promotions) compared with the opt-in framing condition[182] |
| Choice bracketing | Grouping individual choices together into sets | Joint evaluation | Assess multiple candidates simultaneously to promote comparative evaluation | Encourages comparative thinking and reduces reliance on biased heuristics, compared with when candidates are assessed one at a time[183] | In a laboratory experiment, participants who evaluated a male and female candidate simultaneously were less likely to favour men for male-typed tasks (such as mathematics) and women for female-typed tasks (such as verbal tasks), compared with participants who evaluated only a single candidate[184] |
| | | Hiring groups | Make hiring and selection decisions in sets (that is, hire groups of candidates) | Prompts evaluators to consider group composition holistically[218,219] | In a series of online experiments involving real hiring decisions, evaluators constructed more gender-diverse groups when making sets of hiring decisions versus making a single hire ($d$=0.39[a])[185] |
| Pre-commitment | Restricting future options to align future behaviour with goals or values | Pre-commitment to criteria | Evaluators determine and commit to selection criteria before reviewing candidates | Promotes consistency and reduces the risk of justifying biased judgements[190] by anchoring decisions to objective criteria | In laboratory experiments, participants who pre-committed to certain, objective criteria exhibited less tendency to shift standards to fit stereotypes while making selection decisions compared with participants in a control group ($d$=0.40)[191] |
| Linguistic framing | Using language to alter the presentation of information, without changing the underlying choice | Reframing gendered language | Replace gendered language (such as stereotypically masculine words) in organizational communications with more neutral but synonymous alternatives | Signals an inclusive environment and therefore boosts anticipated belonging and application rates among individuals who might feel alienated by gendered language | In laboratory and field experiments, replacing stereotypically masculine language in job advertisements with gender-neutral language increased anticipated belonging and job applications from women and men who are less aligned with masculinity compared with the original job advertisement[193] |
| | | Reframing qualifications more concretely | Specify skills and experiences needed for success using concrete, behaviour-based language | Reduces ambiguity and makes roles feel more accessible to members of under-represented groups who might otherwise consider themselves unqualified | In laboratory and field experiments, clearly defining qualifications using objective and behaviour-based language reduced ambiguity and increased applications from qualified women for high-level jobs ($d$=0.29[a]) compared with using generic language[194] |
| De-identification | Masking identifying characteristics from evaluators | De-identifying application materials | Omit names and demographic information from application materials during evaluation | Removes stereotype and unconscious bias-triggering cues[195] | De-identification mitigated gender bias in experimental hiring tasks[220] and a quasi-natural experiment[196,220] |
| Social norm nudging | Sharing information about what peers or leaders commonly do or believe | Communicate norms that endorse diversity | Have leaders or peers clearly communicate and sincerely endorse attitudes and behaviours that align with diversity and inclusion | Leverages descriptive and injunctive norms to signal what is valued and expected | Classrooms at a North American university shown posters and videos where leaders communicated support for diversity behaved more inclusively than classrooms shown materials where leaders did not communicate support ($d$=0.38[a]) and showed smaller performance disparities between marginalized and non-marginalized students ($d$=0.16[a])[202] |
| Anchoring effects | Shifting the reference or starting point | Rethink rating scales | Use less culturally loaded scale end-points (such as a 6-point scale) to avoid implicit associations between numbers such as 10 or 100 with gendered ideas about brilliance and perfection that tend to favour men | Disrupts biased cognitive anchors and weakens implicit associations[221,222] | A quasi-natural experiment at a North American university showed that using a six-point (versus ten-point) rating scale reduced gender disparities in teaching evaluations in male-dominated fields by limiting the expression of brilliance-related stereotypes[187] |
| Structuring effects | Standardizing procedures and choices by adding structure | Structured interviews | Use standardized interview scripts and scoring rubrics across candidates | Reduces cognitive load, increases consistency and limits reliance on subjective impressions | In an experimental laboratory study, structured interviews reduced selection bias ($r$=0.34) (for example, against overweight candidates) compared with unstructured interviews ($r$=0.34)[101] |

Effect sizes are presented if reported in the original paper or calculable from the published material. $r$, Pearson's $r$ coefficient; $d$, Cohen's $d$ value. [a]The effect size was calculated by the authors of the current Review.

Because each disciplinary approach has distinct assumptions and strengths, the goal is not wholesale adoption of one over the other but a more comprehensive and coordinated approach to designing and implementing strategies for increasing organizational diversity[9].

## Summary and future directions

Organizations are increasingly turning to psychology-based approaches (such as diversity training and 'lean in'-style empowerment programmes) as a primary strategy for promoting diversity. These interventions offer important insights into how individuals think and behave, and can foster short-term changes in attitudes and behaviours. However, when used in isolation, these individual-level solutions are limited in their ability to drive sustained behavioural change or shift important outcomes such as representation, retention and advancement. These limitations stem, in part, from a narrow focus on intrapersonal change and a tendency to overlook the powerful influences of the organizational environments in which decisions are made and behaviour occurs. To overcome these constraints, we propose an integrated behavioural design approach that bridges individual-level and system-level perspectives to offer strategies for redesigning decision environments to support equitable behaviour and outcomes.

In addition to generating practical solutions, a behavioural design perspective can enrich psychology-based approaches by bringing greater attention to contextual influences on behaviour and expanding the focus of behaviour change beyond intrapersonal processes and outcomes. In this way, behavioural design helps to recentre the importance of environment and context on intrapersonal processes and behaviour — core tenets of psychology that have often been overlooked in diversity applications. In turn, psychology can strengthen the emerging field of behavioural design by addressing one of its limitations: a tendency to prioritize real-world impact without fully explaining the mechanisms underlying behaviour change[212]. Incorporating psychological theory can help to ensure that behavioural design interventions are not only effective but also theoretically grounded.

Our Review leads to several recommendations for future research and practice. First, future research should move beyond measuring short-term changes in attitudes or intentions and prioritize evaluating long-term behavioural outcomes such as application rates, selection decisions and actual representation. Moreover, existing studies that test behavioural outcomes tend to focus on the hiring stage and overlook longer-term outcomes such as retention and advancement that might contribute to more organizational diversity at entry levels than at the top, ultimately hindering true progress. Thus, future research should also assess the effectiveness of interventions beyond getting employees through the door to other critical stages of the employee experience. For instance, research should explore how behavioural tools (such as choice architecture, defaults, pre-commitment or social norms) can be applied to promote allyship, reduce bias in performance evaluations or shape more inclusive feedback and decision-making processes. Additionally, future research should explore a wider range of employee experiences beyond representation alone, including how interventions influence employees' sense of belonging, authenticity and thriving at work.

Second, the literature applying behavioural design to organizational diversity is in a relatively nascent stage. Thus, it is still too early to draw firm conclusions about effectiveness, which is why we emphasize its promise and potential only. Future research is needed to expand the empirical evidence for the behavioural design approaches outlined here, as well as to explore novel applications of other behavioural design concepts that have not yet been examined in the context of organizational diversity. Ideally, the evidence base for behavioural design approaches will grow to match what is currently known about psychology-based interventions. As this body of work develops, it is critical that scholars and practitioners apply the same level of empirical scrutiny and rigour that psychology-based approaches have received.

Third, given the context-dependent nature of behaviour, interventions should be tested for external validity and generalizability to the environments where they are intended to be used. Many popular organizational diversity initiatives are drawn from research conducted online or among college students. Laboratory-based research provides valuable foundational insights, but scaling solutions requires validation in real-world organizational settings using field experiments, randomized control trials or quasi-experimental designs. Importantly, journals should support the publication of small or null effects from field studies, as these are essential to building an accurate evidence base that is relevant to and calibrated for practice. Field trials can also inform important design questions[213], such as how to balance efficiency with feasibility, and how to tailor interventions to specific industries, social groups or organizational cultures.

Relatedly, research should investigate why interventions that succeed in the laboratory fail to produce similar results in the field[108,135], with the goal of identifying contextual moderators. For example, field testing in organizations captures how interventions unfold amid social norms, organizational culture and broader societal trends such as diversity, equity and inclusion backlash, factors that are often muted in controlled laboratory settings. In settings with high resistance and backlash towards diversity efforts, behavioural design interventions might be more palatable owing to their relatively 'identity-neutral' framing, which focuses on the features of decision-making environments rather than explicitly naming race or gender bias, and the fact that they preserve employee choice and autonomy rather than imposing a controlling mandate[119]. Still, behavioural design initiatives should be seen as complements to, not replacements for, identity-aware approaches, which remain essential for addressing deeper, systemic inequities[214].

Fourth, practitioners and organizations should engage critically with the empirical evidence behind diversity initiatives rather than defaulting to mainstream but potentially ineffective interventions. Light-touch interventions might seem attractive because they promise large effects for relatively low effort and cost, but they often fail to produce meaningful change. Although more effective interventions require greater investment and more coordination, they are more likely to deliver sustained impact. Researchers and practitioners should collaborate more intentionally to co-create, implement and rigorously evaluate interventions that are grounded in evidence and tailored to specific organizational contexts, and to ensure fidelity in application from theory to practice. More broadly, it is challenging to translate theories from basic science into practice, and intermediary steps (such as field-based empirical testing) are needed before jumping to direct application[215,216]. When done thoughtfully, this translation process not only yields practical insights but also offers opportunities to refine or extend theory. For example, inconsistent results in complex social environments might reveal important boundary conditions.

Finally, we call for a shift away from siloed approaches and towards multi-level strategies that align individual-level interventions with systemic organizational changes. Behavioural design offers a promising framework for designing such organizational-level strategies, enabling organizations to support individual behaviour change and

create environments that make those changes more likely to succeed and last. Importantly, behavioural design should not be viewed as a panacea: although it offers a valuable interdisciplinary perspective that integrates individual and systemic influences on behaviour, solutions that follow from it are likely to inform changes to the more proximal choice environment but might not resolve deeper, systemic inequities. As such, expanding the scope of diversity efforts to include even more macro-level perspectives (such as those from sociology and economics) will be critical for addressing the broader forces that shape inequality in organizations. Future work should explore how to combine interventions across disciplinary perspectives — such as by pairing leadership accountability structures with evaluator nudges or integrating inclusive policy reforms with messaging about social norms — to reinforce and sustain progress.

Published online: 2 October 2025

## References


1. Frith, C. D. Social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 2033–2039 (2008).
2. Fiske, S. T. & Macrae C. N. (eds.) *The SAGE Handbook of Social Cognition* (SAGE, 2012).
3. Neuberg, S. L. & Sng, O. A life history theory of social perception: stereotyping at the intersections of age, sex, ecology (and race). *Soc. Cogn.* **31**, 696–711 (2013).
4. Fiske, S. T. & Neuberg, S. L. in *Advances in Experimental Social Psychology* Vol. 23 (eds. Fiske, S. T. & Neuberg, S. L.) 1–74 (Elsevier, 1990).
5. Brewer, M. B. in *Advances in Social Cognition* Vol. I (eds. Wyer, R. S. Jr & Srull, T. K.) 1–36 (Psychology Press, 2014).
6. Kalev, A., Dobbin, F. & Kelly, E. Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *Am. Sociol. Rev.* **71**, 589–617 (2006).
7. Paluck, E. L., Porat, R., Clark, C. S. & Green, D. P. Prejudice reduction: progress and challenges. *Annu. Rev. Psychol.* **72**, 533–560 (2021).
8. Stephens, N. M., Rivera, L. A. & Townsend, S. S. M. The cycle of workplace bias and how to interrupt it. *Res. Organ. Behav.* **40**, 100137 (2020).
9. Arslan, C., Chang, E. H., Chilazi, S., Bohnet, I. & Hauser, O. P. Behaviorally designed training leads to more diverse hiring. *Science* **387**, 364–366 (2025).
10. Bezrukova, K., Spell, C. S., Perry, J. L. & Jehn, K. A. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychol. Bull.* **142**, 1227–1274 (2016).
11. Allport, G. W. *The Nature of Prejudice* (Addison-Wesley, 1954).
12. Bodenhausen, G. V., Kang, S. K. & Peery, D. in *The SAGE Handbook of Social Cognition* (eds. Fiske, S. T. & Macrae, C. N.) 311–329 (Sage, 2012).
13. Fiske, S. T., Cuddy, A. J. C., Glick, P. & Xu, J. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* **82**, 878–902 (2002).
14. Nicolas, G., Bai, X. & Fiske, S. T. A spontaneous stereotype content model: taxonomy, properties, and prediction. *J. Pers. Soc. Psychol.* **123**, 1243–1263 (2022).
15. Fiske, S. T. Stereotype content: warmth and competence endure. *Curr. Dir. Psychol. Sci.* **27**, 67–73 (2018).
16. Cuddy, A. J. C., Fiske, S. T. & Glick, P. in *Advances in Experimental Social Psychology* Vol. 40 (ed. Zanna, M. P.) 61–149 (Elsevier, 2008).
17. Greenwald, A. G. & Banaji, M. R. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* **102**, 4–27 (1995).
18. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
19. Greenwald, A. G., Nosek, B. A. & Banaji, M. R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197–216 (2003).
20. Kurdi, B. et al. Relationship between the implicit association test and intergroup behavior: a meta-analysis. *Am. Psychol.* **74**, 569–586 (2019).
21. Nosek, B. A. et al. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* **18**, 36–88 (2007).
22. Devine, P. G. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* **56**, 5–18 (1989).
23. Plant, E. & Devine, P. Internal and external motivation to respond without prejudice. *J. Personality Soc. Psychol.* **75**, 811–832 (1998).
24. Dovidio, J. F., Kawakami, K. & Gaertner, S. L. Implicit and explicit prejudice and interracial interaction. *J. Pers. Soc. Psychol.* **82**, 62–68 (2002).
25. Monteith, M. J. Self-regulation of prejudiced responses: implications for progress in prejudice-reduction efforts. *J. Pers. Soc. Psychol.* **65**, 469–485 (1993).
26. Devine, P. G., Monteith, M. J., Zuwerink, J. R. & Elliot, A. J. Prejudice with and without compunction. *J. Pers. Soc. Psychol.* **60**, 817–830 (1991).
27. Smith, E. R. & Zárate, M. A. Exemplar-based model of social judgment. *Psychol. Rev.* **99**, 3–21 (1992).
28. Linville, P. W. The complexity–extremity effect and age-based stereotyping. *J. Pers. Soc. Psychol.* **42**, 193–211 (1982).
29. Linville, P. W. & Fischer, G. W. Exemplar and abstraction models of perceived group variability and stereotypicality. *Soc. Cogn.* **11**, 92–125 (1993).
30. Olson, M. A. & Fazio, R. H. Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Pers. Soc. Psychol. Bull.* **32**, 421–433 (2006).
31. Devine, P. G. & Monteith, M. J. in *Dual-Process Theories in Social Psychology* (eds. Chaiken, S. & Trope, Y.) 339–360 (Guilford, 1999).
32. Greenwald, A. G. et al. A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Rev.* **109**, 3–25 (2002).
33. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford Univ. Press, 1957).
34. Heider, F. Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
35. Kunda, Z. & Spencer, S. J. When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychol. Bull.* **129**, 522–544 (2003).
36. Dovidio, J. F., Gaertner, S. L. & Validzic, A. Intergroup bias: status, differentiation, and a common in-group identity. *J. Pers. Soc. Psychol.* **75**, 109–120 (1998).
37. Tajfel, H. & Turner, J. C. in *The Social Psychology of Intergroup Relations* (eds. Austin, W. G. & Worchel, S.) 33–47 (Brooks/Cole, 1979).
38. Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. *Rediscovering the Social Group: A Self-Categorization Theory* (Basil Blackwell, 1987).
39. Turner, J. C., Oakes, P. J., Haslam, S. A. & McGarty, C. Self and collective: cognition and social context. *Pers. Soc. Psychol. Bull.* **20**, 454–463 (1994).
40. Hogg, M. A. & Terry, D. J. Social identity and self-categorization processes in organizational contexts. *Acad. Manag. Rev.* **25**, 121 (2000).
41. Galinsky, A. D. & Moskowitz, G. B. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *J. Pers. Soc. Psychol.* **78**, 708–724 (2000).
42. Todd, A. R., Bodenhausen, G. V., Richeson, J. A. & Galinsky, A. D. Perspective taking combats automatic expressions of racial bias. *J. Pers. Soc. Psychol.* **100**, 1027–1042 (2011).
43. Todd, A. R., Galinsky, A. D. & Bodenhausen, G. V. Perspective taking undermines stereotype maintenance processes: evidence from social memory, behavior explanation, and information solicitation. *Soc. Cogn.* **30**, 94–108 (2012).
44. Berthold, A., Leicht, C., Methner, N. & Gaum, P. Seeing the world with the eyes of the outgroup—the impact of perspective taking on the prototypicality of the ingroup relative to the outgroup. *J. Exp. Soc. Psychol.* **49**, 1034–1041 (2013).
45. Pettigrew, T. F. Intergroup contact theory. *Annu. Rev. Psychol.* **49**, 65–85 (1998).
46. Dovidio, J. F., Love, A., Schellhaas, F. M. H. & Hewstone, M. Reducing intergroup bias through intergroup contact: twenty years of progress and future directions. *Group. Process. Intergroup Relat.* **20**, 606–620 (2017).
47. Cocco, V. M., Vezzali, L., Stathi, S., Di Bernardo, G. A. & Dovidio, J. F. Mobilizing or sedative effects? A narrative review of the association between intergroup contact and collective action among advantaged and disadvantaged groups. *Pers. Soc. Psychol. Rev.* **28**, 119–180 (2024).
48. Brown, R. & Hewstone, M. in *Advances in Experimental Social Psychology* Vol. 37 (ed. Zanna, M. P.) 255–343 (Elsevier, 2005).
49. Gaertner, S. L. & Dovidio, J. F. *Reducing Intergroup Bias: The Common Ingroup Identity Model* (Psychology Press, 2000).
50. Paluck, E. L., Green, S. A. & Green, D. P. The contact hypothesis re-evaluated. *Behav. Public. Policy* **3**, 129–158 (2019).
51. Gurin, P., Nagda, B. A. & Zuniga, X. *Dialogue Across Difference: Practice, Theory, and Research on Intergroup Dialogue* (Russell Sage Foundation, 2013).
52. Locke, E. A. & Latham, G. P. *A Theory of Goal Setting & Task Performance* (Prentice-Hall, 1990).
53. Locke, E. A. & Latham, G. P. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *Am. Psychol.* **57**, 705–717 (2002).
54. Locke, E. A. & Latham, G. P. *Goal Setting: A Motivational Technique That Works!* (Prentice-Hall, 1984).
55. Wood, R. E., Mento, A. J. & Locke, E. A. Task complexity as a moderator of goal effects: a meta-analysis. *J. Appl. Psychol.* **72**, 416–425 (1987).
56. Ajzen, I. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **50**, 179–211 (1991).
57. Gollwitzer, P. M. & Sheeran, P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. *Adv. Exp. Soc. Psychol.* **38**, 69–119 (2006).
58. Gollwitzer, P. M. Implementation intentions: strong effects of simple plans. *Am. Psychol.* **54**, 493–503 (1999).
59. Fishbein, M. & Ajzen, I. *Predicting and Changing Behavior: The Reasoned Action Approach* (Psychology Press, 2010).
60. Bandura, A. Self-efficacy mechanism in human agency. *Am. Psychol.* **37**, 122 (1982).
61. Wood, R. & Bandura, A. Social cognitive theory of organizational management. *Acad. Manage. Rev.* **14**, 361 (1989).
62. Gist, M. E. & Mitchell, T. R. Self-efficacy: a theoretical analysis of its determinants and malleability. *Acad. Manage. Rev.* **17**, 183 (1992).
63. Locke, E. A., Frederick, E., Lee, C. & Bobko, P. Effect of self-efficacy, goals, and task strategies on task performance. *J. Appl. Psychol.* **69**, 241–251 (1984).
64. Festinger, L. Cognitive Dissonance. *Sci. Am.* **207**, 93–106 (1962).
65. Caza, B. B., Vough, H. & Puranik, H. Identity work in organizations and occupations: definitions, theories, and pathways forward. *J. Organ. Behav.* **39**, 889–910 (2018).
66. Ellemers, N., Spears, R. & Doosje, B. Self and social identity. *Annu. Rev. Psychol.* **53**, 161–186 (2002).
67. Ellemers, N. & Haslam, S. A. in *Handbook of Theories of Social Psychology* (eds. van Lange, P. A. M., Kruglanski, A. & Higgins, E. T.) 379–398 (SAGE, 2012).
68. Steele, C. M., Spencer, S. J. & Aronson, J. Contending with group image: the psychology of stereotype and social identity threat. *Adv. Exp. Soc. Psychol.* **34**, 379–440 (2002).

69. Cohen, G. L. & Garcia, J. Identity, belonging, and achievement: a model, interventions, implications. *Curr. Dir. Psychol. Sci.* **17**, 365–369 (2008).
70. Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R. & Crosby, J. R. Social identity contingencies: how diversity cues signal threat or safety for African Americans in mainstream institutions. *J. Pers. Soc. Psychol.* **94**, 615–630 (2008).
71. Cheryan, S., Plaut, V. C., Davies, P. G. & Steele, C. M. Ambient belonging: how stereotypical cues impact gender participation in computer science. *J. Pers. Soc. Psychol.* **97**, 1045–1060 (2009).
72. Good, C., Rattan, A. & Dweck, C. S. Why do women opt out? Sense of belonging and women's representation in mathematics. *J. Pers. Soc. Psychol.* **102**, 700–717 (2012).
73. Walton, G. M. & Cohen, G. L. A question of belonging: race, social fit, and achievement. *J. Pers. Soc. Psychol.* **92**, 82–96 (2007).
74. Baumeister, R. F. & Leary, M. R. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* **117**, 497–529 (1995).
75. Martens, A., Johns, M., Greenberg, J. & Schimel, J. Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. *J. Exp. Soc. Psychol.* **42**, 236–243 (2006).
76. Steele, C. M. & Aronson, J. Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* **69**, 797–811 (1995).
77. Inzlicht, M. & Kang, S. K. Stereotype threat spillover: how coping with threats to social identity affects aggression, eating, decision making, and attention. *J. Pers. Soc. Psychol.* **99**, 467–481 (2010).
78. Napp, C. & Breda, T. The stereotype that girls lack talent: a worldwide investigation. *Sci. Adv.* **8**, eabm3689 (2022).
79. Wood, W. & Eagly, A. H. A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. *Psychol. Bull.* **128**, 699–727 (2002).
80. Eagly, A. H. & Steffen, V. J. Gender stereotypes stem from the distribution of women and men into social roles. *J. Pers. Soc. Psychol.* **46**, 735–754 (1984).
81. Hirsh, J. B. & Kang, S. K. Mechanisms of identity conflict: uncertainty, anxiety, and the behavioral inhibition system. *Pers. Soc. Psychol. Rev.* **20**, 223–244 (2016).
82. Kang, S. K. & Bodenhausen, G. V. Multiple identities in social perception and interaction: challenges and opportunities. *Annu. Rev. Psychol.* **66**, 547–574 (2015).
83. Heilman, M. E. Sex bias in work settings: the lack of fit model. *Res. Organ. Behav.* **5**, 269–298 (1983).
84. Eagly, A. H. & Karau, S. J. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* **109**, 573–598 (2002).
85. Stephens, N. M., Townsend, S. S. M. & Dittmann, A. G. Social-class disparities in higher education and professional workplaces: the role of cultural mismatch. *Curr. Dir. Psychol. Sci.* **28**, 67–73 (2019).
86. Sandberg, S. *Lean In: Women, Work, and the Will to Lead* (Random House, 2013).
87. Chrobot-Mason, D., Hoobler, J. & Burno, J. Lean in versus the literature: an evidence-based examination. *Acad. Manag. Perspect.* **33**, 110–130 (2019).
88. Cohen, G. L. & Sherman, D. K. The psychology of change: self-affirmation and social psychological intervention. *Annu. Rev. Psychol.* **65**, 333–371 (2014).
89. Sherman, D. K. & Cohen, G. L. The psychology of self-defense: self-affirmation theory. *Adv. Exp. Soc. Psychol.* **38**, 183–242 (2006).
90. Steele, C. M. The psychology of self-affirmation: sustaining the integrity of the self. *Adv. Exp. Soc. Psych.* **21**, 261–302 (1988).
91. Trope, Y. & Neter, E. Reconciling competing motives in self-evaluation: the role of self-control in feedback seeking. *J. Pers. Soc. Psychol.* **66**, 646–657 (1994).
92. Spencer, S. J., Steele, C. M. & Quinn, D. M. Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* **35**, 4–28 (1999).
93. Bandura, A. *Social Learning Theory* (Prentice Hall, 1977).
94. Lockwood, P. & Kunda, Z. Superstars and me: predicting the impact of role models on the self. *J. Pers. Soc. Psychol.* **73**, 91–103 (1997).
95. Lockwood, P. "Someone like me can be successful": do college students need same-gender role models? *Psychol. Women Q.* **30**, 36–46 (2006).
96. Cialdini, R. B., Kallgren, C. A. & Reno, R. R. in *Advances in Experimental Social Psychology* Vol. 24 (ed. Zanna, M. P.) 201–234 (Academic, 1991).
97. Sherif, M. *The Psychology of Social Norms* (Harper, 1936).
98. Bicchieri, C. *The Grammar of Society: The Nature and Dynamics of Social Norms* 1–260 (Cambridge Univ. Press, 2005).
99. Cialdini, R. B. & Goldstein, N. J. Social influence: compliance and conformity. *Annu. Rev. Psychol.* **55**, 591–621 (2004).
100. Cialdini, R. B., Reno, R. R. & Kallgren, C. A. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* **58**, 1015–1026 (1990).
101. O'Reilly, C. A. & Chatman, J. A. Culture as social control: corporations, cults, and commitment. *Res. Organ. Behav.* **18**, 157–200 (1996).
102. Koenig, A. M. & Eagly, A. H. Evidence for the social role theory of stereotype content: observations of groups' roles shape stereotypes. *J. Pers. Soc. Psychol.* **107**, 371–392 (2014).
103. Devine, P. G. & Ash, T. L. Diversity training goals, limitations, and promise: a review of the multidisciplinary literature. *Annu. Rev. Psychol.* **77**, 403–429 (2021).
104. FitzGerald, C., Martin, A., Berner, D. & Hurst, S. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol.* **7**, 29 (2019).
105. Kalinoski, Z. T. et al. A meta-analytic evaluation of diversity training outcomes: diversity training. *J. Organ. Behav.* **34**, 1076–1104 (2013).
106. Nishii, L. H., Khattab, J., Shemla, M. & Paluch, R. M. A multi-level process model for understanding diversity practice effectiveness. *Acad. Manag. Ann.* **12**, 37–82 (2018).
107. Lai, C. K. et al. Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* **143**, 1765–1785 (2014).
108. Costa, E. Examining the effectiveness of interventions to reduce discriminatory behavior at work: an attitude dimension consistency perspective. *J. Appl. Psychol.* https://doi.org/10.1037/apl0001215 (2024).
109. Chang, E. H. et al. The mixed effects of online diversity training. *Proc. Natl Acad. Sci. USA* **116**, 7778–7783 (2019).
110. Lai, C. K. et al. Reducing implicit racial preferences: II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* **145**, 1001–1016 (2016).
111. Homan, A. C., Buengeler, C., Eckhoff, R. A., van Ginkel, W. P. & Voelpel, S. C. The interplay of diversity training and diversity beliefs on team creativity in nationality diverse teams. *J. Appl. Psychol.* **100**, 1456–1467 (2015).
112. Leslie, L. M. Diversity initiative effectiveness: a typological theory of unintended consequences. *Acad. Manage. Rev.* **44**, 538–563 (2019).
113. Caleo, S. & Heilman, M. E. What could go wrong? Some unintended consequences of gender bias interventions. *Arch. Sci. Psychol.* **7**, 71–80 (2019).
114. Dover, T. L., Kaiser, C. R. & Major, B. Mixed signals: the unintended effects of diversity initiatives. *Soc. Issues Policy Rev.* **14**, 152–181 (2020).
115. Castilla, E. J. & Benard, S. The paradox of meritocracy in organizations. *Adm. Sci. Q.* **55**, 543–676 (2010).
116. Kaiser, C. R. et al. Presumed fair: ironic effects of organizational diversity structures. *J. Pers. Soc. Psychol.* **104**, 504–519 (2012).
117. Duguid, M. M. & Thomas-Hunt, M. C. Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes. *J. Appl. Psychol.* **100**, 343–359 (2015).
118. Dover, T. L., Major, B. & Kaiser, C. R. Members of high-status groups are threatened by pro-diversity organizational messages. *J. Exp. Soc. Psychol.* **61**, 1–6 (2015).
119. Dobbin, F., Schrage, D. & Kalev, A. Rage against the iron cage: the varied effects of bureaucratic personnel reforms on diversity. *Am. Sociol. Rev.* **80**, 1014–1044 (2015).
120. Ku, G., Wang, C. S. & Galinsky, A. D. The promise and perversity of perspective-taking in organizations. *Res. Organ. Behav.* **35**, 79–102 (2015).
121. Longmire, N. H. & Harrison, D. A. Seeing their side versus feeling their pain: differential consequences of perspective-taking and empathy at work. *J. Appl. Psychol.* **103**, 894–915 (2018).
122. Lemmer, G. & Wagner, U. Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions: meta-analysis of contact interventions. *Eur. J. Soc. Psychol.* **45**, 152–168 (2015).
123. Saguy, T., Tausch, N., Dovidio, J. F. & Pratto, F. The irony of harmony: intergroup contact can produce false expectations for equality. *Psychol. Sci.* **20**, 114–121 (2009).
124. Mendoza, S. A., Gollwitzer, P. M. & Amodio, D. M. Reducing the expression of implicit stereotypes: reflexive control through implementation intentions. *Pers. Soc. Psychol. Bull.* **36**, 512–523 (2010).
125. Stewart, B. D. & Payne, B. K. Bringing automatic stereotyping under control: implementation intentions as efficient means of thought control. *Pers. Soc. Psychol. Bull.* **34**, 1332–1345 (2008).
126. Mendoza, S., Shorinko, J., Martin, S. & Martone, L. The effects of perspective taking implementing intentions on employee evaluations and hostile sexism. *Pers. Assess. Decis.* **5**, 7 (2019).
127. Holland, R. W., Aarts, H. & Langendam, D. Breaking and creating habits on the working floor: a field-experiment on the power of implementation intentions. *J. Exp. Soc. Psychol.* **42**, 776–783 (2006).
128. Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. Using implementation intentions prompts to enhance influenza vaccination rates. *Proc. Natl Acad. Sci. USA* **108**, 10415–10420 (2011).
129. Smith, B. D. & Silk, K. Cultural competence clinic: an online, interactive, simulation for working effectively with Arab American Muslim patients. *Acad. Psychiatry* **35**, 312–316 (2011).
130. Kuntz, J. C. & Searle, F. Does bystander intervention training work? When employee intentions and organisational barriers collide. *J. Interpers. Violence* **38**, 2934–2956 (2023).
131. Jennings, L., Zhao, K., Faulkner, N. & Smith, L. Mapping bystander intervention to workplace inclusion: a scoping review. *Hum. Resour. Manag. Rev.* **34**, 101017 (2024).
132. Haynes-Baratz, M. C., Metinyurt, T., Li, Y. L., Gonzales, J. & Bond, M. A. Bystander training for faculty: a promising approach to tackling microaggressions in the academy. *N. Ideas Psychol.* **63**, 100882 (2021).
133. Coker, A. L. et al. RCT testing bystander effectiveness to reduce violence. *Am. J. Prev. Med.* **52**, 566–578 (2017).
134. Fenwick, K. M. et al. Challenges to addressing patient-perpetrated sexual harassment in veterans affairs healthcare settings. *J. Gen. Intern. Med.* **36**, 2332–2338 (2021).
135. Mobasseri, S., Srivastava, S. B. & Kray, L. J. A brief social-belonging intervention in the workplace: evidence from a field experiment. *Acad. Manag. Discov.* **7**, 85–103 (2021).
136. Martin, A. E. & Phillips, K. W. What "blindness" to gender differences helps women see and do: implications for confidence, agency, and action in male-dominated environments. *Organ. Behav. Hum. Decis. Process.* **142**, 28–44 (2017).
137. Burnette, J. L., Pollack, J. M. & Hoyt, C. L. Individual differences in implicit theories of leadership ability and self-efficacy: predicting responses to stereotype threat. *J. Lead. Stud.* **3**, 46–56 (2010).

138. Murphy, M. C., Steele, C. M. & Gross, J. J. Signaling threat: how situational cues affect women in math, science, and engineering settings. *Psychol. Sci.* **18**, 879–885 (2007).
139. Hoyt, C. L. & Murphy, S. E. Managing to clear the air: stereotype threat, women, and leadership. *Leadersh. Q.* **27**, 387–399 (2016).
140. Kinias, Z. & Sim, J. Facilitating women's success in business: interrupting the process of stereotype threat through affirmation of personal values. *J. Appl. Psychol.* **101**, 1585–1597 (2016).
141. Cislaghi, B. et al. Changing social norms: the importance of "organized diffusion" for scaling up community health promotion and women empowerment interventions. *Prev. Sci.* **20**, 936–946 (2019).
142. Rudman, L. A. & Glick, P. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* **57**, 743–762 (2001).
143. Rudman, L. A. & Phelan, J. E. Backlash effects for disconfirming gender stereotypes in organizations. *Res. Organ. Behav.* **28**, 61–79 (2008).
144. Heilman, M. E. Description and prescription: how gender stereotypes prevent women's ascent up the organizational ladder. *J. Soc. Issues* **57**, 657–674 (2001).
145. He, J. C. & Kang, S. K. Covering in cover letters: gender and self-presentation in job applications. *Acad. Manage. J.* **64**, 1097–1126 (2021).
146. Kim, J. Y., Fitzsimons, G. M. & Kay, A. C. Lean in messages increase attributions of women's responsibility for gender inequality. *J. Pers. Soc. Psychol.* **115**, 974–1001 (2018).
147. Dennehy, T. C. & Dasgupta, N. Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proc. Natl Acad. Sci. USA* **114**, 5964–5969 (2017).
148. James, E. H. Race-related differences in promotions and support: underlying effects of human and social capital. *Organ. Sci.* **11**, 493–508 (2000).
149. Ragins, B. R. & Cornwell, J. M. Pink triangles: antecedents and consequences of perceived workplace discrimination against gay and lesbian employees. *J. Appl. Psychol.* **86**, 1244–1261 (2001).
150. Welbourne, T. M., Rolf, S. & Schlachter, S. The case for employee resource groups: a review and social identity theory-based research agenda. *Pers. Rev.* **46**, 1816–1834 (2017).
151. Forscher, P. S. et al. A meta-analysis of procedures to change implicit measures. *J. Pers. Soc. Psychol.* **117**, 522–559 (2019).
152. Inzlicht, M., Schmeichel, B. J. & Macrae, C. N. Why self-control seems (but may not be) limited. *Trends Cogn. Sci.* **18**, 127–133 (2014).
153. Asch, S. Opinions and social pressure. *Nature* **176**, 1009–1011 (1955).
154. Asch, S. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr. Gen. Appl.* **70**, 1–70 (1956).
155. Bond, R. & Smith, P. B. Culture and conformity: a meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychol. Bull.* **119**, 111–137 (1996).
156. Acker, J. Hierarchies, jobs, bodies: a theory of gendered organizations. *Gend. Soc.* **4**, 139–158 (1990).
157. Cheryan, S. & Markus, H. R. Masculine defaults: identifying and mitigating hidden cultural biases. *Psychol. Rev.* **127**, 1022–1052 (2020).
158. Acker, J. Inequality regimes. *Gend. Soc.* **20**, 441–464 (2006).
159. Hsu, N., Badura, K. L., Newman, D. A. & Speach, M. E. P. Gender, "masculinity," and "femininity": a meta-analytic review of gender differences in agency and communion. *Psychol. Bull.* **147**, 987–1011 (2021).
160. Diekman, A. B. & Schmader, T. in *The Oxford Handbook of Social Cognition* (eds. Carlston, D. E., Hugenberg, K. & Johnson, K. L.) 471–501 (Oxford Univ. Press, 2024).
161. Prentice, D. A. & Carranza, E. What women should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes. *Psychol. Women Q.* **26**, 269–281 (2002).
162. Paluck, E. L. & Green, D. P. Prejudice reduction: what works? A review and assessment of research and practice. *Annu. Rev. Psychol.* **60**, 339–367 (2009).
163. He, J. C. & Kang, S. K. Moving from i-frame to s-frame focus in equity, diversity, and inclusion research, practice, and policy. *Behav. Brain Sci.* **46**, e159 (2023).
164. Leslie, L. M., Mayer, D. M. & Kravitz, D. A. The stigma of affirmative action: a stereotyping-based theory and meta-analytic test of the consequences for performance. *Acad. Manage. J.* **57**, 964–989 (2014).
165. Thaler, R. H. & Sunstein, C. R. *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale Univ. Press, 2008).
166. Dai, H. et al. Behavioural nudges increase COVID-19 vaccinations. *Nature* **597**, 404–409 (2021).
167. Milkman, K. L. et al. A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proc. Natl Acad. Sci. USA* **119**, e2115126119 (2022).
168. Beshears, J., Dai, H., Milkman, K. L. & Benartzi, S. Using fresh starts to nudge increased retirement savings. *Organ. Behav. Hum. Decis. Process.* **167**, 72–87 (2021).
169. Hallsworth, M., List, J. A., Metcalfe, R. D. & Vlaev, I. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. *J. Public. Econ.* **148**, 14–31 (2017).
170. Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *Am. Econ. Rev.* **104**, 3003–3037 (2014).
171. Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E. & Galinsky, A. D. The critical role of second-order normative beliefs in predicting energy conservation. *Nat. Hum. Behav.* **2**, 757–764 (2018).
172. Balafoutas, L. & Sutter, M. Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* **335**, 579–582 (2012).
173. Leibbrandt, A., Wang, L. C. & Foo, C. Gender quotas, competitions, and peer review: experimental evidence on the backlash against women. *Manage. Sci.* **64**, 3501–3516 (2018).
174. Cialdini, R. B. & Trost, M. R. in *The Handbook of Social Psychology* (eds. Gilbert, D. T., Fiske, S. T. & Lindzey, G.) 151–192 (McGraw-Hill, 1998).
175. Niederle, M. & Vesterlund, L. Do women shy away from competition? Do men compete too much? *Q. J. Econ.* **122**, 1067–1101 (2007).
176. Niederle, M. & Vesterlund, L. Gender and competition. *Annu. Rev. Econom.* **3**, 601–630 (2011).
177. Rudman, L. A., Johnson, K., Julian, S., Phillips, E. & Zehren, K. Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *J. Pers. Soc. Psychol.* **74**, 629–645 (1998).
178. Jachimowicz, J. M., Duncan, S., Weber, E. U. & Johnson, E. J. When and why defaults influence decisions: a meta-analysis of default effects. *Behav. Public. Policy* **3**, 159–186 (2019).
179. McKenzie, C. R. M., Liersch, M. J. & Finkelstein, S. R. Recommendations implicit in policy defaults. *Psychol. Sci.* **17**, 414–420 (2006).
180. Johnson, E. J., Bellman, S. & Lohse, G. L. Defaults, framing and privacy: why opting in-opting out. *Mark. Lett.* **13**, 5–15 (2002).
181. Johnson, E. J. & Goldstein, D. Do defaults save lives? *Science* **302**, 1338–1339 (2003).
182. He, J. C., Kang, S. K. & Lacetera, N. Opt-out choice framing attenuates gender differences in the decision to compete in the laboratory and in the field. *Proc. Natl Acad. Sci. USA* **118**, e2108337118 (2021).
183. Bazerman, M. H. & Moore, D. A. *Judgment in Managerial Decision Making* (Wiley, 2013).
184. Bohnet, I., van Geen, A. & Bazerman, M. When performance trumps gender bias: joint vs. separate evaluation. *Manage. Sci.* **62**, 1225–1234 (2016).
185. Chang, E. H., Kirgios, E. L., Rai, A. & Milkman, K. L. The isolated choice effect and its implications for gender diversity in organizations. *Manage. Sci.* **66**, 2291–2799 (2020).
186. Rivera, L. A. & Tilcsik, A. Scaling down inequality: rating scales, gender bias, and the architecture of evaluation. *Am. Sociol. Rev.* **84**, 248–274 (2019).
187. Kutcher, E. J. & Bragger, J. D. Selection interviews of overweight job applicants: can structure reduce the bias? *J. Appl. Soc. Psychol.* **34**, 1993–2022 (2004).
188. Kivetz, R. & Simonson, I. Self-control for the righteous: toward a theory of precommitment to indulgence. *J. Consum. Res.* **29**, 199–217 (2002).
189. Milkman, K. L., Rogers, T. & Bazerman, M. H. Harnessing our inner angels and demons: what we have learned about want/should conflicts and how that knowledge can help us reduce short-sighted decision making. *Perspect. Psychol. Sci.* **3**, 324–338 (2008).
190. Bryan, G., Karlan, D. & Nelson, S. Commitment devices. *Annu. Rev. Econom.* **2**, 671–698 (2010).
191. Uhlmann, E. L. & Cohen, G. L. Constructed criteria: redefining merit to justify discrimination. *Psychol. Sci.* **16**, 474–480 (2005).
192. Chang, L. W. & Cikara, M. Social decoys: leveraging choice architecture to alter social preferences. *J. Pers. Soc. Psychol.* **115**, 206–223 (2018).
193. He, J. C. & Kang, S. K. De-biasing job ads by replacing masculine language increases gender diversity of applicant pools. *Proc. Natl Acad. Sci. USA* **122**, e2409854122 (2025).
194. Coffman, K. B., Collis, M. R. & Kulkarni, L. Whether to apply. *Manage. Sci.* **70**, 4649–4669 (2024).
195. Bertrand, M. & Mullainathan, S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* **94**, 991–1013 (2004).
196. Goldin, C. & Rouse, C. Orchestrating impartiality: the impact of "blind" auditions on female musicians. *Am. Econ. Rev.* **90**, 715–741 (2000).
197. Behaghel, L., Crépon, B. & Barbanchon, T. L. Do anonymous resumes make the field more even? Evidence from a randomized field experiment. https://api.semanticscholar.org/CorpusID:27320451 (2012).
198. Chang, L. W. & Chang, E. H. On the limits of anonymization for promoting diversity in organizations. *Pers. Soc. Psychol. Bull.* https://doi.org/10.1177/01461672241304593 (2025).
199. Behaghel, L., Crépon, B. & Le Barbanchon, T. Unintended effects of anonymous résumés. *Am. Econ. J. Appl. Econ.* **7**, 1–27 (2015).
200. Roy, E. et al. A contest study to reduce attractiveness-based discrimination in social judgment. *J. Pers. Soc. Psychol.* **128**, 508–535 (2025).
201. Goldstein, N. J., Cialdini, R. B. & Griskevicius, V. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *J. Consum. Res.* **35**, 472–482 (2008).
202. Murrar, S., Campbell, M. R. & Brauer, M. Exposure to peers' pro-diversity attitudes increases inclusion and reduces the achievement gap. *Nat. Hum. Behav.* **4**, 889–897 (2020).
203. Avilova, T. & Goldin, C. What can UWE do for economics? *AEA Pap. Proc.* **108**, 186–190 (2018).
204. Cervantez, J. A. & Milkman, K. L. Can nudges be leveraged to enhance diversity in organizations? A systematic review. *Curr. Opin. Psychol.* **60**, 101874 (2024).
205. Axt, J. & To, J. How can debiasing research aid efforts to reduce discrimination? *Pers. Soc. Psychol. Rev.* **29**, 81–105 (2024).
206. Mertens, S., Herberz, M., Hahnel, U. J. J. & Brosch, T. The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl Acad. Sci. USA* **119**, e2107346118 (2022).
207. Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A. & Aczel, B. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. *J. Behav. Decis. Mak.* **31**, 355–366 (2018).
208. Milkman, K. L. et al. Megastudies improve the impact of applied behavioural science. *Nature* **600**, 478–483 (2021).
209. DellaVigna, S. & Linos, E. RCTs to scale: comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).

210. Hummel, D. & Maedche, A. How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies. *J. Behav. Exp. Econ.* **80**, 47–58 (2019).
211. Chater, N. & Loewenstein, G. The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray. *Behav. Brain Sci.* **46**, 1–60 (2022).
212. Krijnen, J. M. T., Tannenbaum, D. & Fox, C. R. Choice architecture 2.0: behavioral policy as an implicit social interaction. *Behav. Sci. Policy* **3**, 1–18 (2017).
213. Jachimowicz, J. M. Embracing field studies as a tool for learning. *Nat. Rev. Psychol.* **1**, 249–250 (2022).
214. Ely, R. J. & Meyerson, D. E. Advancing gender equity in organizations: the challenge and importance of maintaining a gender narrative. *Organization* **7**, 589–608 (2000).
215. IJzerman, H. et al. Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
216. List, J. A. Optimally generate policy-based evidence before scaling. *Nature* **626**, 491–499 (2024).
217. Dinner, I., Johnson, E. J., Goldstein, D. G. & Liu, K. Partitioning default effects: why people choose not to choose. *J. Exp. Psychol. Appl.* **17**, 332–341 (2011).
218. Read, D., Loewenstein, G. & Rabin, M. Choice bracketing. *J. Risk Uncertain.* **19**, 171–197 (1999).
219. Bordalo, P., Gennaioli, N. & Shleifer, A. Salience theory of choice under risk. *Q. J. Econ.* **127**, 1243–1285 (2012).
220. Fath, S., Larrick, R. P. & Soll, J. B. Blinding curiosity: exploring preferences for "blinding" one's own judgment. *Organ. Behav. Hum. Decis. Process.* **170**, 104135 (2022).
221. Tversky, A. & Kahneman, D. Judgment under uncertainty: heuristics and biases. *Science* **185**, 1124–1131 (1974).
222. Chapman, G. B. & Johnson, E. J. in *Heuristics and Biases* 120–138 (Cambridge Univ. Press, 2002).
223. Berdahl, J. L., Cooper, M., Glick, P., Livingston, R. W. & Williams, J. C. Work as a masculinity contest. *J. Soc. Issues* **74**, 422–448 (2018).
224. Hall, W., Schmader, T., Inness, M. & Croft, E. Climate change: an increase in norms for inclusion predicts greater fit and commitment for women in STEM. *Group. Process. Intergroup Relat.* **25**, 1781–1796 (2022).
225. Cheryan, S., Lombard, E. J., Hailu, F., Pham, L. N. H. & Weltzien, K. Global patterns of gender disparities in STEM and explanations for their persistence. *Nat. Rev. Psychol.* **4**, 6–19 (2024).
226. Glick, P., Berdahl, J. L. & Alonso, N. M. Development and validation of the masculinity contest culture scale: development and validation of the masculinity contest culture scale. *J. Soc. Issues* **74**, 449–476 (2018).
227. Vial, A. C., Muradoglu, M., Newman, G. E. & Cimpian, A. An emphasis on brilliance fosters masculinity-contest cultures. *Psychol. Sci.* **33**, 595–612 (2022).
228. Bian, L., Leslie, S. J., Murphy, M. C. & Cimpian, A. Messages about brilliance undermine women's interest in educational and professional opportunities. *J. Exp. Soc. Psychol.* **76**, 404–420 (2018).
229. Dobbin, F., Kim, S. & Kalev, A. You can't always get what you need: organizational determinants of diversity programs. *Am. Sociol. Rev.* **76**, 386–411 (2011).
230. Iyer, A. Understanding advantaged groups' opposition to diversity, equity, and inclusion (DEI) policies: the role of perceived threat. *Soc. Personal. Psychol. Compass* https://doi.org/10.1111/spc3.12666 (2022).
231. Nishii, L. H. The benefits of climate for inclusion for gender-diverse groups. *AMJ* **56**, 1754–1774 (2013).
232. Hall, W., Schmader, T., Aday, A., Inness, M. & Croft, E. Climate control: the relationship between social identity threat and cues to an identity-safe culture. *J. Pers. Soc. Psychol.* **115**, 446–467 (2018).
233. Hamedani, M. G., Markus, H. R., Hetey, R. C. & Eberhardt, J. L. We built this culture (so we can change it): seven principles for intentional culture change. *Am. Psychol.* **79**, 384–402 (2024).
234. Cheryan, S. & Muragishi, G. A. Removing masculine defaults in the hiring process. *Proc. Natl Acad. Sci. USA* **122**, e2501630122 (2025).



## Acknowledgements
The authors thank the Institute for Gender and the Economy (GATE) and the Behavioral Economics in Action at Rotman (BEAR) research centres at the University of Toronto Rotman School of Management, as well as the Social Sciences and Humanities Research Council of Canada (SSHRC), for their support.


## Author contributions
The authors contributed equally to all aspects of the article.

## Competing interests
The authors declare no competing interests.

## Additional information
**Peer review information** *Nature Reviews Psychology* thanks Aneeta Rattan, Andrea Vial, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.