All inside our heads?

Möller, Christian; Passam, Saffron; Riley, Sarah; Robson, Martine

Published in:
Gender, Work, and Organization

DOI:
10.1111/gwao.13028

Publication date:
2023

Citation for published version (APA):
Möller, C., Passam, S., Riley, S., & Robson, M. (2023). All inside our heads? A critical discursive review of unconscious bias training in the sciences. *Gender, Work, and Organization*. Advance online publication. https://doi.org/10.1111/gwao.13028

Received: 9 May 2023 | Accepted: 15 May 2023

DOI: 10.1111/gwao.13028

WILEY

ORIGINAL ARTICLE

Gender and Race in the International Sciences: Organizational Practices of Diversity

All inside our heads? A critical discursive review of unconscious bias training in the sciences

Christian Möller[1] | Saffron Passam[2] | Sarah Riley[3] | Martine Robson[2]

[1]School of Social Work & Social Policy, University of Strathclyde, Glasgow, UK

[2]Department of Psychology, Aberystwyth University, Aberystwyth, UK

[3]School of Psychology, Massey University, Palmerston North, New Zealand

Correspondence

Christian Möller, School of Social Work & Social Policy, Lord Hope Building, University of Strathclyde, 141 St James Road, Glasgow G4 0LT, UK.
Email: christian.moller@strath.ac.uk

Funding information

Engineering and Physical Sciences Research Council, Grant/Award Number: EP/S011927/1

Abstract

In response to persistent systemic gendered and racial exclusions in the sciences, unconscious or implicit bias training is now widely established as an organizational intervention in Higher Education (HE). Recent systematic reviews have considered the efficacy of unconscious bias training (UBT) but not the wider characteristics and effects of the interventions themselves. Guided by feminist scholarship in critical psychology and post-structuralist discourse theory, this article critically examines UBT across STEMM and in HE institutions with a discursive analysis of published studies. Drawn from systematic searches in 4 databases, we identify three types of UBT reported in 22 studies with considerable variation in intervention types, target groups, and evaluation methods. Guided by limited cognitive problematizations of unconscious bias as a problem located inside individual minds, interventions follow established patterns in neoliberal governmentality and make available specific feeling rules and subject positions. These current Equality, Diversity & Inclusion practices present a new technology of power through which organizations may regulate affect and behavior but leave structural inequalities and barriers to inclusion intact.

KEYWORDS
equality and diversity, governmentality, implicit bias, psychologization, unconscious bias

1 | INTRODUCTION

This article reviews publications relating to unconscious bias training (UBT) interventions across STEMM (Science, technology, engineering, mathematics, and medicine) areas in Higher Education (HE) institutions. A growing number of interventions and evaluations try to address historic and structurally ingrained forms of racism, sexism, and other discrimination from a positivist paradigm, which locates the source of the problem in psychological biases and provides solutions in measurable testing and creative training courses. Recent systematic reviews have considered the efficacy of UBT, but not the wider characteristics and effects of the interventions themselves or their implications for how we think about the effects of racism and sexism. Guided by feminist scholarship in critical psychology and post-structuralist discourse theory, our discursive approach makes visible how UBT situates the problem of and solution to organizational bias at the level of the individual. In a novel approach, we draw upon Hochschild's (1979, 2015) "feeling rules" to examine how a new EDI (Equality, Diversity & Inclusion) industry may reproduce knowledge, power relations, and affects that work against enacting wider change at a societal, organizational, and structural level. Feeling rules point to the patterned yet invisible social systems of management that result in a transmutation, or change, in the private ways that we use feelings (Hochschild, 2015). Critical engagement with the underlying discourses and power effects highlights the role of individualizing practices in the perpetuation of othering and prejudice with the apportioning of blame and responsibility. We counter a dominant empiricist reductionism in intervention evaluation with a feminist focus on knowledge creation and ownership to open a much-needed space for critical reflection on marketization and psychologization of gendered and racialized inequalities in the sciences.

STEMM departments have responded to EDI concerns through a multitude of interventions seeking to support underrepresented groups where interventions often claim a range of benefits (Marshall et al., 2020), mainly focusing on training opportunities and improving individual career outcomes. In sharp contrast to these "success stories" and contradicting the myth of the inclusive and post-racial university (Sian, 2019), recent research has highlighted persistent racism and racial inequality in academia (Bhopal & Pitkin, 2020; Dupree & Boykin, 2021) alongside gender disparities (Casad et al., 2021), which are then seen as the outcome of social discrimination and histories or racialized violence and exclusion. Although there are differences in discriminatory experiences of women, people who are Black and/or from underrepresented racial and ethnic backgrounds, and people with disabilities, all negotiate the dominant norm of the white, male, heteronormative, able-bodied scientist (Ahmed, 2004; Liu, 2017). This positioning problematizes the availability/accessibility of affirmative subjectivities for underrepresented members of the STEMM community. In particular, "doing inclusion" is interrogated through a questioning of the inclusive subject and conditions of possibility that are formed by the construction of an inclusive subject (Brewis, 2019) and the complex interconnection of people, organizations, and society, and intersectionalities (Rodriguez et al., 2016). The reach of inclusivity work has therefore stretched from the question of what an organization is doing and toward a critical integration of what it is that the doing is doing. As Ahmed (2012) outlines, diversity work is often nonperformative, meaning that statements of commitment do not necessarily produce the effect that they name, and thus do not improve the conditions of possibility for the people that diversity work aims to support. In this sense, action is not necessarily additive, and change does not necessarily address the problem.

Existing reviews of UBT seek to evaluate its effectiveness. We argue that this can perpetuate empiricist concerns, producing a particular perspective on "what works?". In contrast, the present article maps out and critically reflects on the power relations and consequences of the individualizing and psychologizing discourses that underpin UBT. We also consider how the literature in turn has constitutive material effects for forming new spaces and subjectivities in HE. Drawn from a systematic search in four electronic databases, we first summarize the scope and reported

outcomes across 22 studies in STEMM and HE settings. We thereby extend the traditional scoping review with an added concern for epistemic reflexivity and problematizations, that is, how a social problem in need of intervention is constructed and thereby rendered governable (Triantafillou, 2012) with material effects. Our analysis theorizes the ways in which resistance to intervention is understood in terms of the "science" of social cognition and the authority of expert psychologists. Building on and substantially extending existing critiques of unconscious bias, we show how current EDI practices are not merely limited in their effectiveness, but present a new technology of power through which organizations may regulate affect and behavior. We further explore individual behavior change and affect management as familiar technologies of neoliberal governmentality. Our discussion then expands on established critiques of unconscious bias research to highlight its psychologizing function and cultivation of confessional rituals that create a spectacle of EDI work, which leaves structural inequalities and institutional power relations untouched. Finally, drawing on intersectional perspectives and critical feminist psychology, we propose new lines of inquiry which require better epistemic reflexivity and a sensitivity for the role of psychology in naturalizing and reproducing gendered and racialized inequalities in STEMM and wider society.

2 | BACKGROUND

2.1 | Gender, underrepresentation, and exclusions in STEMM

Despite some positive trends and top-level commitments for investment in increasing diversity (CASE, 2014; Guyan & Oloyede, 2019; The Royal Society, 2021; Xiao et al., 2020), the lack of diversity across STEMM stubbornly persists. Identification of the problem of underrepresentation is visible throughout strategic and operational frameworks (see, e.g., the (UKRI, 2021) equality and diversity inclusion strategy). Advance HE (2020) publishes an annual HE data report, covering all academic, professional, and support staff within UK HEIs. In 2020, the academic workforce of Chemistry and Physics were 92% white, while 71% and 80% were male, respectively. Historically, much of the criticism of a lack of diversity in STEMM has focused on access or lack thereof. The argument is that fewer people from diverse backgrounds entering STEMM means fewer people progress through the organization and into leadership roles. However, interventions that result in increases in diversity in younger groups do not automatically result in gains further upstream, reflected in metaphors, such as the leaky pipe, glass ceiling, and conveyor belt. Terms now embedded into organizational language to signify the EDI problem indicate a complex combination of factors that reduce the possibilities for women and people from underrepresented backgrounds to thrive in academic careers.

A key way to understand the social production of discrimination is through the notion of an ideal worker, especially potent for STEMM given competitive, individualistic, and solitary norms and stereotypes of the successful white male scientist (Ong et al., 2018). Fagan and Teasdale (2021) point to the presence of gendered substructures within STEMM, norms, and behaviors that naturalize expected performativity, such as networking and flexibility, but which are exclusionary and the product of particular power relations. Women scientists are, for example, less likely to hold more prestigious roles at conferences (Johnston et al., 2016) or achieve recognition awards (Silver et al., 2018), both of which provide the recipient with a substantial contribution to their academic profile and assimilate the performance of excellence synonymous with the successful scientist. Other social roles intersect with the ideal worker and contribute to the devaluing of talent. Forty-three percent of women leave full-time STEM employment after the birth of their first child, compared to 23% of men (Cech & Blair-Loy, 2019), supporting lived-experience accounts whereby motherhood and professional legitimacy are constructed as mutually exclusive (Herman et al., 2013; Thébaud & Taylor, 2021).

The reasons for a lack of diversity in STEMM are multiple, but a key proposition is that women and underrepresented groups struggle to access opportunities equitably at recruitment, promotion, and progression stages. In terms of early-career research fellowships, in 2019, 39% of offers were made to women compared to 61% men, 8% of BAME people compared to 92% non-BAME, and 1% of people with a declared disability compared with 99% who

do not (The Royal Society, 2021). Linguistic bias at important progression gateways is shown to contribute to the production of discrimination. Moss-Racusin et al. (2012) found that both male and female STEMM academics rated male candidates as significantly more competent and hirable than their female counterparts, offering higher salaries and mentoring to men. More broadly in academia, women's achievements are also shown to be evaluated more negatively than male candidates', with personnel responsible for selection shown to use linguistic selection bias to maintain male power, hinder women's careers, and build invisible barriers (Rubini & Menegatti, 2014).

Even when women and people from underrepresented groups employ male linguistic norms synonymous with the ideal worker identity, they run the risk of being perceived as aggressive and confrontational, thus conforming to stereotypical tropes, like the "angry Black woman" (McGee & Bentley, 2017), but may also need to distance themselves from characteristics deemed to be feminine, such as "taking things personally" (Rhoton, 2011). Experiences of discrimination and prejudice can lead to lower emotional well-being, fewer cognitive resources, and lower performance under the pressure of stereotypes (Carlone & Johnson, 2007; Gutiérrez y Muhs et al., 2012; Mavin & Williams, 2013). Together, these issues deflect from institutional responsibility for systematic disadvantage and impact on individuals by reducing their sense of belonging in STEMM and further fracturing their capacity to occupy valued subjectivities. Women and underrepresented groups in STEMM therefore experience discrimination, either explicit, unintentional, or unconscious, despite anti-discrimination legislation and a cultural valuing of equality (Acker, 2006; Riley, 2002).

STEMM has responded to EDI concerns through a multitude of interventions at the individual level, including training for career development (Chang et al., 2016), leadership (Bickel et al., 2002), research training (Byars-Winston et al., 2011), and psychosocial support through mentoring, sponsorship, and coaching (Farkas et al., 2019; Huston et al., 2019; Lewis et al., 2016). These reflect the popularity of individual-level interventions that claim to respond to the problem of a lack of representation/diversity in STEMM, while also aligning with neoliberal discourses of individual responsibility. Self-help (Riley et al., 2019), for example, is particularly instrumental, whereby the ideal worker readily uptakes personalized problematizations on behalf of the institution and is consequently positioned as the agent of change. Indeed, evaluations of interventions often paint a positive picture of success, claiming a range of benefits, including the cultivation of a sense of solidarity, belonging, or connectedness (Marshall et al., 2020).

The rising organizational attention and business case for a diverse workforce coincided with a shift in discourse from equality toward inclusion. Consequently, there has been a proliferation of corporate consultancies offering deliverable training packages for organizations that "manage diversity" and support the realization of worker value (Oswick & Noon, 2014). However, individual-level interventions that are disconnected from the social production of discrimination are unlikely to be successful from an anti-discrimination perspective. Programs have been criticized for taking a "fixing the women" approach, for example, (Rottenberg, 2014). This individualized focus, which may offer an expedient "quick fix," often leaves the power structures responsible for inequities and exclusions unchallenged (Laver et al., 2018; van den Brink and Stobbe, 2014). Measures to address the lack of diversity in contemporary HEI through leadership programs, such as the Race Equality Charter and Athena Swan, have also been critiqued for circumventing structural change and perpetuating white privilege and feminisms adaptable to market environments (Tzanakou & Pearce, 2019). Thus, while those delivering interventions often claim success, there are a range of critiques, particularly within organizational studies, that call for a shift from "business-case" friendly solutions that commodify difference, toward praxis, or the integration of theory and practice (Pullen et al., 2017).

This landscape contributes to understandings of academia as an "ivory tower," a space that retains undertones of whiteness, classism, and other exclusions (Souto-Manning & Ray, 2007). A lack of diversity has far-reaching implications, including the loss of the benefits of diversity, resulting in deficiencies in attention paid to the gendered dimension of science and a lack of gender balance in decision-making (Rees, 2001). Together, gray literature, research on biases in STEMM, and a plethora of lived experience accounts (Ong et al., 2018) indicate the presence of bias and point toward need for intervention. As a result, researchers have sought to explore bias from various vantage points, pushing beyond the simplistic premise that fewer diverse people enter and flourish in science because there is less diversity in the pipeline.

2.2 | Unconscious bias research and critiques

There exist multiple mainstream definitions of bias. A popular definition from Greenwald and Banaji (1995, p. 4) proposes implicit/unconscious bias as "past experience influences judgment in a fashion not introspectively known," thus implying an unreflective pattern of thinking that influences the behavior of the person. The theoretical constitution of unconscious bias is multifaceted, but broadly draws upon attitudes, stereotypes, and experience that a person "holds." Unconscious bias relates to schematic categorizations that involve "like me/not like me" evaluations/identifications that are thought to permeate sexism, racism, and other forms of prejudice. Consequentially, a range of thinking, actions, and outcomes are understood, explained, and made sense of through a lens of unconscious bias.

The recognition of the problem of bias, whatever the understanding of the underlying nature and causes, has produced a range of interventions that aim to increase understanding, expectations, and/or competence through a popularized notion of bias literacy (Sevo & Chubin, 2010). By far the most established of diversity interventions is UBT which draws upon the well-established psychological constructs of implicit prejudices and social preferences believed to be outside of conscious awareness and control. The approach was popularized through the development of the Implicit Association Test (IAT), first introduced in 1998 by Greenwald et al. (1998), and now readily available through online tests such as "Project Implicit" (2021) at Harvard University. The IAT measures implicit attitudes by comparing response times between associations and evaluations, such as tendencies for associating Black faces with negative words. Participants quickly sort words on a computer screen by pressing appropriate keys for the "good" or "bad" category in a series of rounds where concepts are switched around. Following the IAT, participants are given a score indicating the degree of their implicit preference for some group. Since its introduction, the methodology of the IAT and its evidence base have been heavily contested and in December 2020 the UK Cabinet Office (2020) announced all UBT would be phased out from the Civil Service, citing lack of evidence base following a review by the Government's Behavioral Insights Unit.

Nevertheless, UBT remains by far the most widely established form of EDI training in HE settings (Equality Challenge Unit, 2013), where it is often delivered through mandatory online courses. The operationalization of UBT can be EDI specific, but also subsumed into other administrative functions, such as Research Excellence Framework reviewer training (University of Nottingham, 2022). Despite its popularity, several systematic reviews have contested the effectiveness of interventions aimed at reducing unconscious bias and stereotypes in organizations. A comprehensive evidence review by the Equality and Human Rights Commission (Atewologun et al., 2018) provided a thorough assessment of the evidence base across 18 studies, concluding that despite some evidence of raising awareness and reducing implicit bias, there was little evidence showing that interventions lead to effective behavior change. The review also highlighted potential dangers of "back-firing" effects due to the highlighting of negative stereotypes and various evidence gaps, notably the exclusion of perspectives from people with protected characteristics.

A key study within the review from Lai et al. (2014) found that 9 out of 17 assessed interventions were ineffective and that no intervention led to a lasting reduction in explicit racial preferences. A recent meta-analysis by Calanchini et al. (2020) concluded that half the included IAT interventions did not change associations, only two had any over more than a few days, and that some even increased associations between black faces and negative words. Another systematic review by Fitzgerald et al. (2019) of 30 studies found that reported effects varied considerably by intervention type, with bias-reduction strategies and exposure to counter-stereotypical exemplars among the more effective interventions, while perspective taking was shown to be least successful. However, due to small sample sizes and lack of rigorous evidence across many studies, the review concludes that "many interventions have no effect or may even increase implicit biases" (Fitzgerald et al., 2019, p. 10) and recommends that any training be integrated into sustained in-depth programs with a view to bringing about structural and organizational changes.

Both Noon (2018) and Kahn (2018) point to a conservative political agenda behind the "color blindness" among UB interventions where racism is said to be a thing of the past, replaced by the concept of ever-present and empirically visible cognitive biases. *Blind Spot*, Banaji and Greenwald's (2016) highly influential book, is widely cited within the UB literature and by proponents of the IAT. The authors argue that "explicit bias is infrequent; implicit bias is

pervasive" and are careful to assert that findings from the IAT do not indicate racism, but merely bias. In addition to ignoring the persistent legacies of racism and colonial histories, the promotion of bias testing and awareness training in "post-racial" societies ignores the diverse forms of racist practices where "symbolic, modern and color-blind racists are aware of their biases and do not conceal their views since they are expressing a socially acceptable form of racism" (Noon, 2018, p. 201). A key problem with the use of the IAT as a predictor for future behavior is not just the assumed color blindness where people should no longer see race and respond to Black and White faces in the same way (Kahn, 2018), but also that where racist beliefs may be explicit and believed to be socially acceptable, people are unlikely to be swayed by awareness training. For Noon (2018), participants' actual willingness to change remains a crucial but unaddressed issue, together with the continued disregard for how organizational structures are reproducing biases and acting as obstacles to well-intentioned individual acts.

3 | GOVERNING EDI AT A DISTANCE

In previous sections, we have shown how UBT and other interventions deployed in the name of EDI have failed to address historical deficits around underrepresentation and the persistent structural determinants around gendered and racialized worker norms and unequal access to resources. For example, scientists who conform to dominant norms have access to resources and skills that enable them to negotiate academia more successfully than those who struggle to be recognized as legitimately belonging (Kahn & Ginther, 2017). From this perspective, EDI terminology, such as "BAME," helps to construct and reproduce differences, enabling the othering of groups in relation to a white, male scientist norm. Normalization is important, since it is the process by which social constructions of race, gender, and class come to be understood as "natural" and universal, which legitimizes the categorization and treatment of particular groups (Frambach & Martimianakis, 2017). Townley's (1997) turn to institutional isomorphism sought to explain how material practices are utilitarian and intended to increase homogeneity in teams that then work to legitimize the organizational aims and objectives. The situated reproduction of norms then deserves much more critical attention as these offer orientation for appropriate conduct within organizations.

We have further shown that fundamental critiques of the IAT and its disregard for wider structural causes are not new. However, despite often calling out the individualizing effects of neoliberal policies and their impact on HE, these perspectives often remain limited by pointing only to what is disregarded (structural change). This still neglects the productive effects of neoliberal discourses and how individualizing power operates: Subjectivity itself becomes the primary site of intervention and constant problematization of internal capacities (Chandler & Reid, 2016). Indeed, Foucault maintained consistent concerns for the production of subjectivity and historically situated "modes of subjectification" operating at a micro level within organizations (Lemke, 2011, p. 172) acting on bodies and creating new moral forms of being and new identities as workers. A critical discursive perspective further challenges any easy separation between these objects and processes of knowledge production (Beetz & Schwab, 2017) and any claims to value freedom or distanced and objective truth claims. We here draw on governmentality studies (Dean, 2010; Lemke, 2019) and Foucault's (1991, p. 176) broad understanding of government as "the set of institutions and practices, from administration to education, through which people's conduct is guided" to make visible the workings of power, and possible resistances, in EDI interventions and accompanying research processes. Power is then always productive and "operates on the field of possibilities in which the behavior of active subjects is able to inscribe itself" (Foucault, 2002, p. 341) at all levels of an institution. Following Rose (2001, p. 6), the concept of governmentality brings attention to how the "conduct of individuals is governed 'at a distance', by shaping the ways they understand and enact their own freedom" in academic settings. The key question then becomes *how* individual conduct is guided through EDI interventions, and how these in turn create new knowledge and norms through available feeling rules (see Section 5.4) and subject positions (see Section 5.5).

Feminist psychology, informed by post-structuralist theory, recognizes all knowledge as political, situated, and embedded in gendered and racialized power dynamics. Where traditional psychology has historically been centered

around individual agency (Parker, 2020a) and conceived of institutional change through modifying individual attitudes, managing affect, and controlling desires, critical feminist psychology remains highly suspicious of individualism and the use of psychology as an interventionist tool. It builds on established critiques of power-knowledge regimes and the psychologization of social justice issues (Vos, 2012) across the psy-disciplines (Rose, 1985, 1998) by "interrogating and problematizing harmful discourses and offering alternative ways of understanding" (Wigginton & Lafrance, 2019, p. 546). Feminist psychology has consistently prioritized questions of positionality and who is creating and holding knowledge, encouraging researchers to "take into account the social institutions that structure public and private life, the political economy, the structures of power and privilege, and cultural ideologies" (Magnusson & Marecek, 2017, p. 26). It locates knowledge production and action not only in formal and distanced research processes but also in moments of "affective dissonance" (Hemmings, 2012), feeling rules (Hochschild, 1979, 2015), and silences (Brown, 2009). In these moments, there is a clear disconnect between expected feelings and affective states guided by institutional discourses on one hand and the socialized power relations with embodied, situated experiences of anger, disgust, or discomfort which are otherwise sanitized and rendered invisible through claims to objective science.

Beyond questions of efficacy and quantifiable measures of success, this critical sensitivity for knowledge production and ownership allows questioning problematizations across interventions, including what aspects of life are considered especially important or problematic, what is deemed irrelevant and marginalized, whose authority as experts is recognized and how truth claims are produced and warranted. The privileging of particular kinds of knowledge over others, such as the value given to science, provides the building blocks for desired subjectivities (Moonesirust & Brown, 2021; Rose, 1998) in a market environment through the indirect targeting of interests, cognitions, decisions, and capacities for choice and individual preferences. Because the psychological subject is at the center of the intervention, the compulsion to build capacities in cultural competence, awareness and self-regulation of biases becomes a source of human capital (Lorenzini, 2018), which is never fulfilled but requires constant self-optimization. Kahn (2018) here points to the origins of the IAT in behavioral economics, which according to McMahon (2015) has become a new technology of power used to depoliticize social, political, and economic interventions by making them appear as purely technical, grounded in science and ultimately benign. Consequently, we must recognize academic EDI research itself as a valuation practice which constantly produces and disseminates knowledge as discursive capital within institutions where some subject positions absorb value (Angermuller, 2018), while others remain marginalized.

Overall, this theoretical perspective challenges positivist approaches to EDI work by drawing attention to the productive effects of its discursive work, and what is at stake when becoming objects of EDI interventions and new neoliberal subjects. It also brings attention to how difficult emotions can be contained, carefully managed, and translated into more "positive" ways while resistance is minimized. Our analysis in the following sections contributes to these conceptual debates by exploring the productive effects of UBT, which make it into a conduit for neoliberal governmentality (Martin & Waring, 2018). Research here becomes a vehicle for producing new knowledge about the problem of gendered and racial marginalization in HEI, constructing not only the problem and solutions as inherently psychological, but also offering new subject positions within this framework. With these come new discursive limits (Lemke, 2011) to the sayable and thinkable, but also affective–discursive practices (Wetherell, 2012) shaping the conditions of possibility for how we might feel about injustices if we are to collectively move beyond current forms of discrimination.

4 | METHOD

Our search strategy included systematic electronic searches in 4 databases (Psycinfo, Scopus, ERIC, Education Abstracts), complemented by manual searches in select journals and following reference lists from included articles. For database searches, we combined basic keywords of unconscious or implicit bias with (AND) the area of interest,

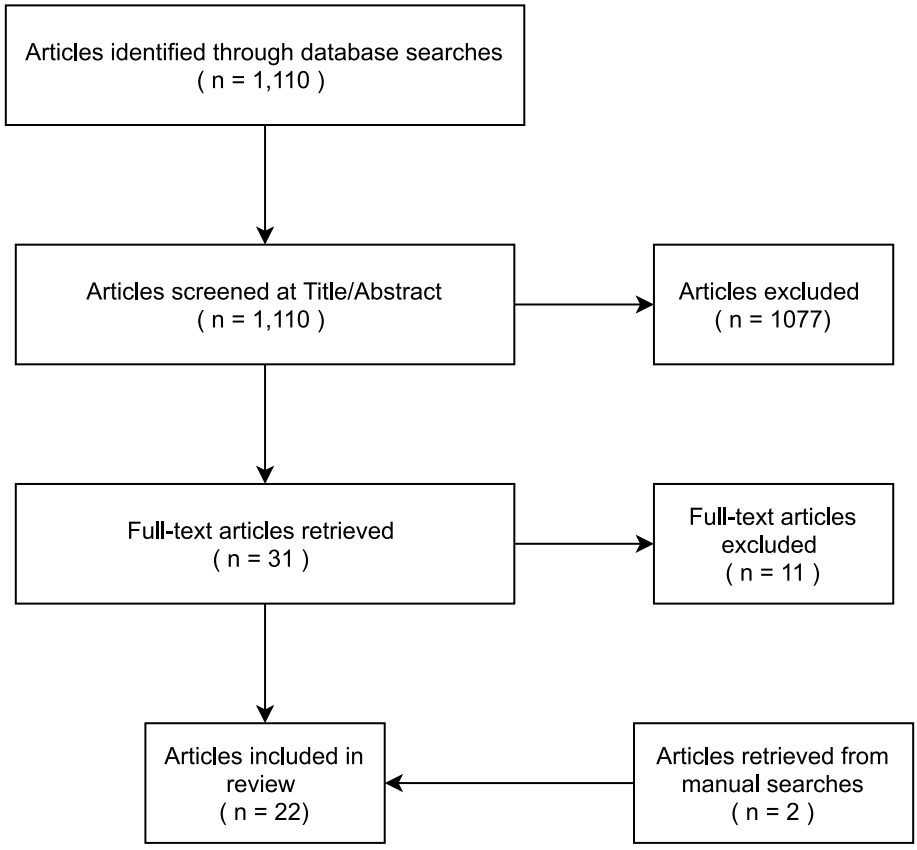

FIGURE 1 Study selection flow chart.

that is, STEMM (including medicine), science, university, and (OR) common terms like intervention, course, or training. With 1110 hits across four databases, all references were downloaded and initially screened for inclusion at abstract and title. To be included, articles had to document some intervention addressing unconscious bias in STEMM areas or HE settings and provide some evidence or reporting of outcomes, results, or effectiveness. We excluded comment and opinion pieces, as well as letters to the editor and studies conducted in health or medicine outside of education. Non-English sources and any populations outside STEMM or HE (e.g., schools) were also excluded. Two additional articles were retrieved through manual searches using reference lists. After removing duplicates and screening 31 full-text articles, a total of 22 studies were included for review (see Figure 1).

Scoping reviews "aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available, and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before" (Mays et al., 2004, p. 194). Arksey and O'Malley (2005) outline four common scenarios where a scoping review might be appropriate, including summarizing and disseminating research findings through detailed descriptions of findings or, alternatively, to map out the extent and nature of research activity in a specific area through a rapid review. Combining these two major concerns, we go beyond detailed mapping and description of studies in the later stages of the review by critically examining the underlying discourses, theoretical assumptions, and subjectifying effects in current interventions. This critical review attempts to map out these discourses, understood as flows of knowledge through time and space (Jäger, 2001) that are informing interventions, and how these interventions in turn have constitutive material effects for forming new spaces and subjectivities in HE.

In practice, we still produced a conventional extraction form to tabulate study characteristics, map out the research landscape, and assess the diversity of methodological approaches (see Tables S1–S3). Rather than identify

gaps in an existing "evidence base," a discursive review here seeks to establish dominant ontological perspectives, asks how the research area comes to be constructed through research activities themselves, and what kind of knowledge is being created in the process. Informed by established approaches in critical discursive psychology (Willig, 2014), the questions guiding our coding and analysis were: What are the ontological positions and epistemological assumptions underpinning interventions? What are the key concepts and values informing the research? What are the limitations and silences imposed by these assumptions and what questions remain unasked? How are participants being positioned within the research and what subject positions are made available? Finally, what constitutes a successful intervention and what are the wider social implications?

Electronic versions of all articles were imported into MAXQDA (v. 20) for coding and analysis. Guided by the review questions, we coded problematizations of unconscious bias (e.g., definitions, related concepts, and key literature), intervention details (e.g., aims and objectives, target populations, activities, and outcomes), and implications for subjectivity (e.g., participants' quotes on behavior change, affect, and resistance). The coding was an iterative process and some subcodes were added and articles later recoded. We used the memo function to note observations and add key summaries for sections which were then linked to codes for later retrieval. As in traditional scoping reviews (Arksey & O'Malley, 2005), we created templates and then populated extraction tables by using coding queries and tabulated these by intervention types (see Tables S1–S3) to compare intervention aims, setting, activities, concepts and theories, and reported outcomes. Guided by Wetherell's work on affective–discursive practices (Calder-Dawe et al., 2021; Wetherell, 2012), we analyzed how these constructions of bias created available affective–discursive positions and with them new feeling rules (Hochschild, 1979) and normative expectations. For Hochschild (2015), feeling rules can be identified by "inspecting how we assess our feelings, how other people assess our emotional display, and by sanctions issuing from ourselves and from them." Consequently, we revisited all coded subject positions to look for desirable attitudes, reactions, or affective responses and quickly noted common patterns around the management of blame and any "bad feelings" linked to the intervention.

5 | FINDINGS

Whereas traditional scoping reviews often report findings by themes; below we first provide an overview of interventions before identifying and critically examining the main discourses and theories informing their design, delivery, and evaluation.

5.1 | The scope of unconscious bias training in HE

Of the 22 interventions we identified, 7 were conducted in academic STEMM contexts, 7 were conducted in health education (including medical schools) and 8 with other university students or staff. Participant numbers varied considerably ranging between 16 and 450 (see Tables S1–S3) and were not always reported. Furthermore, reporting of demographic information was sporadic and not meaningful to outline here as a group. Studies frequently reported attendance figures without further breakdown by participants and those who went on to complete the course and any evaluation measures. Some studies did not distinguish between intervention sites and participants, leaving unclear who exactly the target of the intervention was, and raising concerns about a lack of consistency and rigor.

Despite considerable variability in intervention types, course contents and modalities of delivery, we identified three broad types of UBT: (1) workshops and learning activities delivered and assessed through IATs, (2) more interactive workshops with creative formats, and (3) teaching-focused diversity and bias literacy training sessions. Although intervention aims and activities frequently overlap, we find it useful to distinguish these three groups to illustrate the UBT landscape. In the following sections, we summarize intervention aims, problematizations, activities, and documented outcomes before turning to our own critically informed evaluation to outline implications for institutional EDI cultures.

Seven out of the 22 included studies used some form of IAT to assess the efficacy of UBT (see Table S1). Employing either experimental or before–after designs, these studies administered IATs at baseline, post-intervention, and sometimes at longitudinal follow-up (Devine et al., 2012) to report reductions in unconscious bias. Notably, the IAT was not only used as a validated instrument to assess intervention effects, but its delivery made up the core part of the intervention itself. These courses aimed to make participants better aware of their unconscious biases and their impact on decision-making in the organization. The IAT exercises here took a crucial role as an "unconsciousness raising tool" (Casad et al., 2012) to make visible and confront participants with their unconscious biases. Not all studies conducted an IAT as a baseline measure before the learning activities but instead invited participants to reflect on the IAT itself (Casad et al., 2012) or briefly introduced them to psychological literature on unconscious bias before taking the IAT (Ghoshal et al., 2013). Activities and presentations of the nature of unconscious bias varied between combinations of scientific definitions, interactive slideshows (Carnes et al., 2012), videos and articles (Gatewood et al., 2019), and reflective writing and group exercises (Cahn, 2017).

Five of the 22 interventions entailed more interactive workshops and creative learning activities, centered around group discussions and reflexive learning (see Table S2). While Hannah and Carpenter-song (2013) combined introductory course readings on unconscious bias with reflective writing exercises, participants in Lueke and Gibson's (2016) completed a mindfulness exercise, and West et al. (2019) used visual mapping exercises to make visible the diversity in participants' own networks. Saetermoe et al. (2017) report on a wider empowerment and mentoring program, which combines readings, field trips, and social activities together with more formal courses in critical race theory. Rather than targeting individual staff, the focus here was on learning about and overcoming institutional barriers and transforming institutional research cultures.

Finally, 10 of the 22 included studies provided some form of training in bias and diversity literacy with a stronger emphasis on knowledge transfer and incorporating course contents into staff training and the curriculum (see Table S3). Although both Carnes et al. (2012) and Devine et al. (2017) conducted IATs before the workshops, the IAT was used more as an introductory activity followed by presentations and group discussions around "habit breaking" strategies. Similarly, Adams, III et al. (2014) report administering IATs before and after the workshop but do not evaluate and report the findings in relation to teaching activities. Online courses (Hutchins & Goldstein Hode, 2019) set out to convey the value of diversity to the institution with the aim of increasing self-awareness and cultural competence among university staff. Video presentations (VIDS) were used by Hennes et al. (2018) and Moss-Racusin et al. (2018) on the nature of gender bias with expert interviews followed by strategies of bias reduction. Krutkowski et al. (2019) took a different approach by exploring how media reporting on transgender issues can reinforce unconscious bias where group discussions covered representations in news articles and television.

5.2 | The effectiveness and limitations of UBT

Before turning to more critical theoretical evaluation, we highlight a number of concerns within UBT intervention studies' own paradigm quality criteria. The variability of the UBT in all of its modalities made for a challenging evidence base to interpret and make inferences from. Activities and presentations of the nature of unconscious bias varied between combinations of scientific definitions, interactive slideshows (Carnes et al., 2012), videos and articles (Gatewood et al., 2019), and reflective writing and group exercises (Cahn, 2017). Not all studies conducted an IAT as a baseline measure before the learning activities but instead invited participants to reflect on the IAT itself or briefly introduced them to psychological literature on unconscious bias before taking the IAT (Ghoshal et al., 2013).

Moreover, little attention was paid to the application of psychological theory within the intervention delivery or interdisciplinary engagement. Theory often operated to provide a means to evaluate or legitimize UB training, rather than direct or shape an intervention. Carnes et al. (2015), for example, argue that self-efficacy is a "cornerstone" of behavior change in overcoming gender bias, and Carnes et al. (2012) directly interpret increases in self-efficacy as evidence for the effectiveness of their "gender bias habit-reducing intervention." But it was often taken as a given

that self-efficacy would increase as a result of UBT, neglecting theoretical underpinnings of the construct where self-efficacy is built over time through iterative cycles of failure and success, with appropriate mentorship and in a supportive environment (Bandura, 1988). Furthermore, there was limited acknowledgment of application of psychological theory within the intervention delivery or the macro/micro decisions that may be guided by that theory. Issues are left unresolved, such as whether self-reported intentions to change are better predictors when captured in written form rather than verbal, a strategy employed by Carnes et al. (2015, p. 6). Evidence that human behavior change is difficult (Kelly & Barker, 2016) was largely ignored in constructions of change as being simple to achieve on the basis of attitude change or intention to act: Cahn (2017) remarked that pernicious implicit bias is "easily neutralized," while Devine et al. (2012) state that strategies to alleviate UB would not be "difficult to implement." This overlooks how self-efficacy requires change at several levels, and that interpersonal, institutional, and societal change are clearly beyond the scope of individualized constructs such as personal self-efficacy.

Finally, despite reporting positive feedback and self-perceived changes in awareness, there was limited evidence of how changed attitudes translated into measurable/actual change within the organization/outside of the individual. Three months following the intervention, Carnes et al. (2015) outlined that self-reports of actions to promote gender equity increased significantly when at least 25% of faculty had attended the workshop. While Cahn (2017) reported gains in recruitment from underrepresented groups at one institution, the study design and lack of a control group left it unclear whether these changes were due to the intervention or not. These limitations were also visible in expressions of future directions; Gatewood et al. (2019, 451) recognize a need to "evaluate change in actual bias."

In summary, even from an empirical perspective, the capacity to draw inferences from findings or replicate interventions is extremely limited due to poor study design and lack of rigor in reporting and evaluation. Limited engagement with wider theoretical literature weakened the interventions' capacity to engender change and neglected a substantial evidence base that human behavior is subject to social determinants. Furthermore, by formulating bias in purely psychological terms, there was very limited evidence gathered of how any changes to perception had resulted in changes to wider institutional practices and cultures in the workplace. In the following sections, we develop this analysis through our critical theory lens and examine the discursive work performed by UBT when it psychologizes inequalities, regulates behavior, and manages desirable affective positions.

5.3 | UBT psychologizes inequalities

Different biases were discursively constructed across interventions through dominant, but ultimately limiting, psychological language, which located both the problem and solution of persistent inequalities in HE within the individual mind. The dominant construction of UB was as a bad habit or cognitive error, which essentializes biases as natural psychological phenomena located within the person. The focus on "habits" constructed unconscious bias as undesirable, but everyday occurrence located naturally in the individual which leads to "cognitive error" (Cahn, 2017) if not corrected for. These constructions firmly locate the problem within the person and in their information processing, thereby requiring a psychological tool to fix processing through awareness raising and self-reflection. For example, Devine et al. (2012, p. 1267) state that the "intervention is based on the premise that implicit bias is like a habit that can be reduced through a combination of awareness of implicit bias, concern about the effects of that bias, and the application of strategies to reduce bias." Here, the problem is named as a habit followed by a three-point list of action: awareness, concern, and strategy. Constructing the problem of inequality in the HEI workplace as an unconscious psychological problem positions the solution for UB in psychological work on the self, requiring better awareness, consciousness raising, and reflexivity. Most authors reinforce the notion of UB as "malleable" (Adams, III et al., 2014; Hennes et al., 2018) or constituting a "remediable habit" (Carnes et al., 2012) requiring educational interventions directed at the individual. Workshops and taught courses centered around bias literacy, thereby position participants as being universally vulnerable to UB as a natural condition and require them to actively work on themselves to reduce their biases through better awareness and guided introspection. Thus, although the first step to mitigating

biases was always increasing awareness as initial treatment, UB was also problematized as "cognitive errors" which needed to be "neutralized" through "priming" participants with cultural images contrary to their established beliefs.

The introduction and group teaching of "bias reduction strategies" to "break the prejudice habit" (Devine et al., 2012) are grounded in psychological theories and language of health behavior change which target individuals at the level of subjectivity to bring about the intended changes in self-regulating "thought processes" and social responses (Carnes et al., 2012) and cultivate "cultural competence" as an individual skill set (Hutchins & Goldstein Hode, 2019). Indeed, a range of strategies related to working on the mind were described, and in some cases, understood as being required to target this psychological problem:

> The strategies most frequently employed to counteract personal bias included stereotype replacement, counterstereotype imaging, individuating, and perspective taking. Thus, faculty appeared to be engaging in an intentional integration of bias literacy concepts in their professional lives, prerequisites to changing cultural institutional norms.
>
> (Carnes et al., 2012, p. 73)

We note the importance of wider discourses to supporting the construction of UB and associated remedies. In their attempts to raise awareness and generate reflective discussion, interventions drew on well-established, but narrowly defined notions of unconscious bias from psychological literature. Popular definitions by cognitive psychologists like Greenwald discussed in the literature (Banaji & Greenwald, 1995) were invoked frequently but uncritically. Such grounding of bias in scientific method and experimental evidence was used to demonstrate the negative impact of measured implicit biases on social interaction and health outcomes. Hannah and Carpenter-song (2013), for example, explained that unconscious stereotypes were targeted as cultural "blind spots," "personal backgrounds" which "need to be discovered and minimized to reduce bias." Furthermore, by locating the problem inside the mind, the problem of racism and sexism in HE becomes amenable to the language of health behavior change and expert psychological intervention, mostly via socio-cognitive theories of information processing and associated self-regulation. Devine et al. (2012, p. 3), for example, note the "conceptual parallels" that connect their own intervention with health behavior change and cognitive behavior change. Carnes (2012, p. 6) structured the analysis of their intervention workshops with the popularized Transtheoretical Model (Prochaska & Velicer, 1997), frequently used to direct health interventions such as smoking cessation.

This psychologization worked through related medicalized discourses, whereby UB is constructed as a diagnosable and treatable condition. The habit breaking intervention employed by Devine et al. (2017), for instance, presupposes that overcoming the "mental habit" first requires some diagnostic procedure as part of "becoming aware of when one is vulnerable to unintentional bias." Others advocated the use of IATs as a "consciousness-raising" tool "for individuals who display relatively pronounced implicit biases" (Adams, III et al., 2014, p. 204). Collectively, these discourses reinforce the proposition that without intervention, the individual would remain biased and limited by cognitive constraints, making wider institutional transformation impossible until the individual "stage" had been addressed. In contrast to these psychological deficit-based constructions, Ghoshal et al. (2013) provided a notable exception through an explicit mention of institutional racism. A key finding of our analysis is therefore that the dominant framing of these studies constructed individuals as responsible for, and capable of, changing their behavior around bias, making subjectivity the site of the intervention.

5.4 | Feeling rules

Locating the problem of bias in the individual, and thus subjectivity as the aim of the intervention, had the potential to evoke feelings of culpability, shame, or distress in participants. This logic of blame was often recognized and explicitly disrupted by intervention studies. To understand what flows of affect were understood to be desirable, why and with

what effects, we drew on Hochschild's (1979, 2015) work on how organizations manage human feeling and examined the "feeling rules" of UBT. Our analysis shows how there was an expectation for participants of these intervention studies to feel blame in response to UBT and that these feelings might be required to motivate change. However, these feelings were also refuted through a construction of bias as natural, a dismissal of such feelings as "defensive," and/or the positioning of "bad" feelings as limiting change and thus needing to be moved through quickly. Combined, these feeling rules minimized possibilities for considering the material and affective implications of structural/institutional/systemic racism and sexism.

The expectation for participants of UBT to feel individualized blame formed a key feeling rule and logic of the interventions. For example, Devine et al. (2012) stated that personal concern arises from awareness of being biased; Carnes et al. (2012) described cognitive dissonance as experienced by participants who wanted "to blame the tool rather than me"; while Gatewood et al. (2019, p. 449) reported that participants may be "less receptive" to UBT because of the threat to their professional and egalitarian values. This blame rule was explained in terms of how internalization of racism, sexism, or any form of prejudice is hard to hear, leading to recommendations that interventions needed to first orientate to the participants' feelings. For example, Hutchins and Goldstein Hode (2019, p. 476) report the "willingness to grapple" with feelings of guilt and shame as a step toward cultural competence. Accepting some individual blame in response to UBT and that these feelings might be required to motivate change was therefore a thread that ran through these studies.

However, the need for individuals to shoulder blame was deflected when studies routinely drew on the authority of social–psychological research and cognitive science to construct bias as "natural and ubiquitous rather than a sign of personal failings" (Cahn, 2017, p. 2). Jackson et al. (2014), for example, recommended the use of "non-confrontational" language and inclusive terms to reassure participants that everyone holds these biases. Here, we connect our analysis to critical race studies identifying institutional "double-speak" (Doharty et al., 2021), whereby the need for intervention, via work on the self, simultaneously blames the individual for the bias while also absolving them from the responsibility or blame. Consequently, the person can deflect bad feelings because it is not a conscious decision, but a collective pathology located in the individual. This element of absolving blame was only explicitly recognized once in the dataset where the authors admitted that a narrow focus on UB located inside the individual had resulted in a "'no-blame' discourse" without wider discussion of "any strategies for integrating a micro-sociological approach with more macro-structural factors" (Hannah & Carpenter-song, 2013, p. 335).

Only one study, Ghoshal et al. (2013), significantly questioned the psychological deficit-based model and associated unconscious bias argument. Instead, they argued it was important to convey to participants on intervention programs that "structural racism is not reducible to individual attitudes, whether conscious or unconscious" (Ghoshal et al., 2013, p. 135) and that racial biases were not fully unconscious, but exist on a continuum "anywhere from fully to partially unconscious." Despite their unusual highlighting of institutional racism, these authors still demonstrated a need to deflect defensive feelings by stating that they employed the term UB in the course to make it appear less threatening.

Defensiveness as a problem to be managed was central to our second feeling rule. Several studies included warnings that UBT could provoke defensiveness and related negative feelings that should be planned for, requiring management. Stone et al. (2020), for instance, chose not to share IAT scores with participants prior to the exercises in order to "reduce defensiveness." Other studies reported observations of intervention participants challenging the internal validity of the IAT, which the researchers interpreted as forms of defensiveness (e.g., Gatewood et al., 2019). Such defensiveness was problematized because, while it was constructed as a normal part of UBT, it also required management to ensure continuity of "buy in."

Defensiveness was located as a problem within the individual, for example, as an unwillingness to admit to holding biases (Casad et al., 2012). This meant that wherever the legitimacy of the intervention was questioned by participants, the researchers positioned the participants as defensive and (without irony) defended the concept, closing debate. Researchers drew on a wider discourse of scientism to defend UB, mobilizing "the science" to evidence the interventions' survival over and above interrogation of UB as a concept. Upon recognizing defensiveness, the

response was to draw upon experimental evidence, cognitive models of information processing to elicit a "logical" response to the problem. Gatewood et al. (2019, p. 449) for example, point to the importance of explaining the science in order to deflect disbelief or rejection, while Cahn (2017); points to the "substantial body of evidence supporting the concept". This call to science allowed concerns to be dismissed by falling back to the authority of positivist epistemologies and classic studies in social cognition.

Along with managing blame through the "no blame" of unconsciousness, and understanding defensiveness as a problem in the person to be managed, was a third feeling rule that uncomfortable feelings will happen but need to be moved through quickly. We identify this rule when, for example, Gatewood et al. (2019, p. 448) report the need to move intervention participants from the feelings of defensiveness and guilt toward responsibility. Carnes et al.'s (2012) study reported that one participant refused to take the IAT out of fear of "feeling bad." Similarly, Stone et al. (2020, p. 95) remark that feeling defensive can make someone resistant to change, while Hutchins and Goldstein Hode (2019) explicitly highlight the role of challenging emotions "as obstacles to action" (p. 476), but also as potential "resources for change" that positions challenging feelings as the precursor to action. Although in such studies, feelings of guilt, shame, and defensiveness are understood as necessary motivational drives, in none of these accounts is there a recommendation or expectation for intervention participants to stay with these difficult feelings. The inferred implication of this rule is that bad feelings get in the way of doing the necessary psychological work that is needed, a further form of doublespeak, conveying a message of "feel but stop feeling quickly" or "feel but not in that way." This is despite literature on decolonizing the academy that suggests discomfort needs to be part of reflections on the meanings of whiteness, sexism and privilege (Millner, 2022), and that diversity work includes confrontation and recognition of feelings as an important and legitimate part of the human experience of change (Hunter, 2015).

5.5 | Subject positions

With UBT, psychologization happens at a cognitive and affective level, each contributing to the formation of new subject positions while minimizing the potential for organizational change. Examining the central figures constructed in this body, we identified four subject positions: (i) ideal subject position—the worker/employee with a bad cognitive habit, responsible for and motivated to self-regulate and change under expert guidance. Its corollary is (ii) the abject (Kristeva, 1984): the defensive subject unable to engage in change because they cannot handle the truth. A third subject position is that of the cognitive expert—evaluator, confessor, and absolver; one who cannot be challenged, but is supported by science. Trainers and psychologists designing and delivering interventions remain in privileged positions as expert communicators of scientific knowledge, offering moral absolution from bad feelings, and guiding the "de-biasing" of our minds. There is also a fourth subject position in the agentic individual whose role is to solve the problem of other people's behavior with their own behavior change. In Carnes et al.'s (2012) study, women were being trained to ask for raises—the gap is thus constructed as being in women's skill set rather than in the unequal workplace, and post-feminist discourses take for granted that she will be disadvantaged unless she speaks up (Rottenberg, 2014). Devine et al. (2012) also required individual efforts and supported the idea that individuals affected by institutional sexism, racism, and ableism can make fundamental changes in their organizations provided they had enough skill and motivation.

Such individualized solutions meant that across the included studies, very few engaged meaningfully with the literature on intersectional inequalities that centers structural causes, and even fewer demonstrated how the intervention design had accounted for these, with White-Davis et al. (2018) and Saetermoe et al. (2017) as notable exceptions where interventions themselves were guided by concepts and histories of anti-racist movements and aims beyond merely raising awareness. But rather than providing an analytic framework to understand axes of privilege and oppression, intersectionality was often invoked to rationalize an aspect of complexity or aspect of unrealistic ambition at the individual level (Hennes et al., 2018; West et al., 2019), effectively becoming another tool to encourage "buy-in." Some of the more creative and unstructured courses did include exercises in critical reflexivity,

for example, reading groups engaging with histories of racism (see Hannah & Carpenter-song, 2013), a broader focus on institutional barriers (Saetermoe et al., 2017), and discussion of better support systems (Krutkowski et al., 2019). Yet, it could be argued that these examples offered a shift away from UBT rather than a development of the underlying principles of the intervention per se.

6 | DISCUSSION

Unconscious bias training is an established intervention underpinned by research that evaluates, and often calls into question, its efficacy. The literature landscape includes a broader critique of psychologization in society that highlights how social problems are individualized in ways that draw attention away from social systems, including institutional power and organizational structures and processes; and specifically, critical race studies analyses of EDI interventions. In HE settings, we extend this work by focusing on the research published about UBT interventions, arguing that analyzing this research is important because it is the process by which knowledge is generated. From a critical theory informed perspective, this means that UBT research exercises a form of productive power that renders the issue of gendered and racial marginalization in HEI knowable in a particular way and with this knowledge some actions are legitimized and enabled, while possibilities for others are closed down.

Despite differences across UBT, there were recognizable patterns shared by most of the studies, and thus in this sense they are homogenous. Many of the interventions in our dataset included initial IATs to sensitize people to their biases, followed by pedagogical strategies of building awareness, which varied only in their form of delivery, and all shared conceptions of individual knowledge gaps and lack of awareness. The interventions we analyzed also shared common concerns over eliciting participants' defensiveness, protecting their "buy-in," and even making the training into enjoyable and rewarding experiences. The interventions in our dataset also included problematizations of the efficacy of UBT. Efficacy is extremely difficult to evaluate because, even by the standards of the paradigm in which the interventions were carried out, their evaluation processes were often weak. This finding contributes to the existing literature on efficacy by highlighting the lack of a robust evidence base for UBT. We, however, extend UBT efficacy research significantly by, drawing on Foucault, considering active processes of problematization, and showing that racism and sexism in HE are constructed as located in (i) the individual as the bearer of bias; (ii) in bad feelings that occur when understanding themselves as inherently racist/sexist/prejudiced; and (iii) enabling the problem to be one of fixing a subject. From this, we showed that the solution to the problem of overwhelming numbers of white, able-bodied men in STEMM subjects is constructed as a problem of individual psychology. The implications of these problematizations are threefold, namely that racism/sexism is naturalized, structural critique is bypassed, and as are the difficult feelings that may arise from being the target of blame.

Our analysis also examined how racism and sexism are naturalized in STEMM. UBT takes for granted a racist and sexist environment from which cognitive models and biases have developed and follows traditional social psychology that reduces social processes to measurable individual and group levels. UBT thus only conceives of the possibility for changing environments by changing the individuals that inhabit it, a standpoint which follows the western psychological tradition of rendering invisible the political and cultural context in which knowledge of social phenomena is produced with exclusive concerns for measurable and predictable cognitive attributes (Klein, 2017). Nelson and Zippel (2021) explain the rising popularity of bias training with the privileging of visible "demonstrability," pointing to the key features of relatability, versatility, and impartiality of implicit bias at the individual level. But while demonstrability may mean that an intervention works in an operational sense, opportunity for transformative potential or any disruption is limited in line with liberal discourses of inclusion.

We have shown how naturalization is supported by combining "the science" of UBT with playful activities and relatable, practical exercises that invite institutional participation, but which often trivialize and individualize injustices. Any fundamental challenge to psychological experiments and attempts to expose the structural determinants of inequality, or wider reflections on the meanings of whiteness, sexism, and privilege would be at odds with the

construction of UB as a diagnosable and treatable psychological problem. The narrow emphasis on social cognition as an authoritative science limits the possibility for very different understandings of how and why harm is done. Not only does UBT ignore and absolve individual and collective responsibility for deliberate discriminatory behavior and obscures the benefits of being in privileged positions, it diverts attention from how our everyday working life is shaped by dominant social norms of what constitutes "good" workers and "correct" ways to conduct ourselves as model professionals and successful scientists. By instead naturalizing the internalization of any "bad habits" related to racism and sexism through the concept of UB, any requirement for structural critique is bypassed. The psychologizing effects of the IAT mean that participants are affected by bias only by their shared humanity and natural cognitive defaults, not by their privileged position in society and power held within the institution. The universal nature of the intervention and the IAT also made it flexible enough to include some contextual information on the target demographics and extent of biases in STEMM, but there was no account of what makes universities specifically racist and hostile environments. Therein lies another naive assumption of shared beneficence where all individuals and organizations mean well and will respond positively to complaints of discrimination that may be at odds with lived experiences of, as Ahmed (2010) frames it, becoming "killjoys" by speaking out against injustices.

Along with not recognizing structural discrimination and privilege, our analysis also showed a denial in the value of staying with difficult feelings. Rather than recognize potential meaning and value in emotions and defensive reactions, anger or resistance was quickly discounted. By discounting complex emotions, interventions followed established patterns of neoliberalized management of affect (Adams et al., 2019), which avoids looking at the causes of trauma and violence and only makes visible internal capacities for perpetual optimization and self-regulation, radically abstracted from social and material contexts. In our dataset, we showed that across studies, there were similar appeals to cognitive science, and bias reduction strategies were shown alongside management of the affective in ways that support the work of Tate and Page (2018) outlining that training was more about alleviating feelings of shame through a confessional release of white guilt and experiencing temporary solace.

Against the backdrop of an expanding diversity industry and consultancies in UBT, there have been growing critiques of how EDI interventions effectively psychologize racism and historical inequalities. Ahmed (2012, p. 117) describes how EDI initiatives can become non-performative speech acts, where, for example, personal confessions and discoveries of personal biases are "taken up as if they are performatives (as if they have brought about the effects they name) such that the names come to stand in for the effects." Applied to UBT, rather than a reduction in sexist, racist, and other exclusive and oppressive practices, positive feelings and visible measurements of success constitute a successful non-performance where the IAT and psychological science function and maintain institutional whiteness, a privileged form of knowledge and practice of knowing which discounts racialized experiences and structural inequalities. Building on this analysis, Lee Jackson argues that the psychologizing power of UBT adds a new dimension to this non-performance through the authority of the psy-sciences leaving us resigned "to the 'fact' of racially biased brains" (Lee Jackson, 2018, p. 48). This article has shown how these issues are reproduced in academic scholarship on UB and the reporting of intervention designs.

7 | IMPLICATIONS

Diversity work has been critiqued for its performative and confessional character, promoting a "moderate" feminism (Tzanakou & Pearce, 2019) and offering absolution from white guilt (Tate & Page, 2018) to meet institutional demands for visible and marketable interventions. Unconscious bias training is by far the most established of these interventions, presenting itself as an evidence-based tool capable of detecting and correcting individual biases. Our critical review adds to conceptual critiques of unconscious bias with an original analysis of how different biases are problematized and targeted for intervention in STEMM areas with specific focus on its affective dimensions. We identified three types of UBT training and with considerable variability in content, target, and delivery of interventions. The vast majority of studies was informed by cognitive science and constructed unconscious bias as a cognitive error or bad

habit, amenable to "de-biasing" through awareness training. By reducing all diverse forms of gendered and racialized violence to a cognitive level, we argue that structural sexism and racism and barriers to change remained largely absent. The IAT performed a central function here, both as a key component of interventions themselves and as a supposed predictor of future behavior, ignoring how racist, sexist, and other discriminatory practices are so embedded in institutional cultures that they have become invisible and taken for granted. The diagnosis of UB through the IAT and other introspective activities here worked through confessional practices that absolved individuals of any blame and responsibility for their actions. The potential for harm to the individual or organization caused by the intervention itself was left unaddressed, while the evocation of feelings of discomfort and active resistance were avoided and countered with reference to the "science" underlying the IAT, which further normalizes the status quo.

Meanwhile, the expertise and authority of the trainers remained unquestioned by the researchers, as did the neutrality of psychological "science" and its individualizing effects, the language and power effects of psychology as a form of education (Parker, 2020b). These rituals promise personal growth and development necessary for flourishing in academic spaces and can be repeated as often as desired. A focus on affect management and behavior change was shown to further individualize and psychologize inequalities in line with neoliberal agendas that have had significant impact on how we think about diversity in HE settings. A common limitation of governmentality perspectives is that context-specific and local deviations from dominant norms and available subject positions may be missed. Our critical review was necessarily limited by the accounts selectively reported in the literature where resistance to interventions was managed and marginalized, while the life experience brought by the trainers themselves was largely omitted. This presents an important gap and opportunity for critical research to combine a sensitivity for the workings of neoliberal governmentality with an ethnographic imaginary (Brady, 2016), exploring, for example, the diverse ways that people navigate and develop resistances within institutional EDI cultures.

Yet, UB research acknowledges that bias is not reduced permanently and fills an organizational need in "post-racial" universities for a performance that can be repeated with public visibility to demonstrate commitment to equality and improve prestige without committing to fundamental change (Bhopal, 2020; Sian, 2019). Ahmed (2012) points out that EDI policies can enable racism and sexism to flourish behind tick-box exercises, such as diversity training, and lip service to equality. Extending these broader critiques, we have shown that UBT can function as a spectacle of anti-racism and anti-sexism, which can be consumed as a sanitized and much more palatable version of racial conflict, gendered inequalities, and oppression. Neoliberal governmentality here manages affects by avoiding the intense distress, pain, and violence of racism and sexism, as well as the guilt and shame felt by those who benefit from systemic inequality. Moreover, resistances to UBT remain under-explored, under-theorized, and too easily dismissed by falling back on the authority of positivist epistemologies and classic studies in social cognition. Visible work on the self, rewarded by positive group experiences and personal discoveries of the workings of the mind were accompanied by a duty not to speak about present racism (Hunter, 2015) or overt and conscious discrimination. What remains missing is an appreciation and critical exploration of the perceived threats to identity and potential loss of power when people are asked to reflect on their own role in maintaining barriers to inclusion.

Current debates about UBT inside and outside psychology have focused on its adequacy and effectiveness, often acknowledging that UBT alone is not enough to reduce discrimination (Applebaum, 2019). Here, we see a dangerous fallacy that may further drive the expansion of EDI cultures and marketized interventions without recognizing their own power effects and potential harms. Our analysis has shown that UBT has performative and productive effects in its own right, produces new knowledge of how diversity issues are to be governed at a psychological level, and creates new subjectivities. When deployed uncritically and without engaging with the histories of racism, sexism, and other exclusions in the sciences, we agree that UBT evaluations are not simply insufficient but "insidious" (Applebaum, 2019) in fostering ignorance and reducing structural violence to easily correctable individual habits.

We therefore see much more potential for organizational changes in approaches that foreground structural inequalities, which would include practitioners reflecting critically on their own positionality, who is being targeted, what constitutes a successful intervention, and how it is evidenced (see Table S4 for further suggestions). Intersectional thinking requires acknowledging the dynamic membership of multiple categories and STEMM academics

suffering several structural disadvantages. Secondly, courses should be attentive to diversity and differences between individuals (including neurodiversity) and conceptualize differences as originating from structural inequalities and discursive processes, not naturally given cognitive processes. We have highlighted how researchers can become part of the systems reproducing disadvantage but hope that better understanding of their location within these systems of power will enable, facilitate, and encourage critical reflexivity toward a more equitable future.

ACKNOWLEDGMENTS

This work was supported by Engineering and Physical Sciences Research Council (EPSRC)/Grant number EP/S011927/1.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Christian Möller https://orcid.org/0000-0003-1731-5431
Sarah Riley https://orcid.org/0000-0001-6712-6976

REFERENCES

Acker, Joan. 2006. "Inequality Regimes: Gender, Class, and Race in Organizations." *Gender & Society* 20(4): 441–64. https://doi.org/10.1177/0891243206289499.

Adams, Glenn, Sara Estrada-Villalta, Daniel Sullivan, and Hazel Rose Markus. 2019. "The Psychology of Neoliberalism and the Neoliberalism of Psychology." *Journal of Social Issues* 75(1): 189–216. https://doi.org/10.1111/josi.12305.

Adams, Virgil H., III, Thierry Devos, Luis M. Rivera, Heather Smith, and Luis A. Vega. 2014. "Teaching about Implicit Prejudices and Stereotypes: A Pedagogical Demonstration." *Teaching of Psychology* 41(3): 204–12. https://doi.org/10.1177/0098628314537969.

Advance HE. 2020. "Equality+ Higher Education: Staff Statistical Report 2020." https://www.advance-he.ac.uk/media/5941.

Ahmed, Sara. 2004. "Declarations of Whiteness: The Non-Performativity of Anti-Racism." *Borderlands* 3(2). https://webarchive.nla.gov.au/awa/20050616083826/http://www.borderlandsejournal.adelaide.edu.au/vol3no2_2004/ahmed_declarations.htm.

Ahmed, Sara. 2010. "Happy Objects." In *The Affect Theory Reader*, edited by Melissa Gregg and Gregory J. Seigworth. Durham, NC: Duke University Press.

Ahmed, Sara. 2012. *On Being Included: Racism and Diversity in Institutional Life*. Durham, NC: Duke University Press.

Angermuller, Johannes. 2018. "Accumulating Discursive Capital, Valuating Subject Positions. From Marx to Foucault." *Critical Discourse Studies* 15(4): 414–25. https://doi.org/10.1080/17405904.2018.1457551.

Applebaum, Barbara. 2019. "Remediating Campus Climate: Implicit Bias Training is Not Enough." *Studies in Philosophy and Education* 38(2): 129–41. https://doi.org/10.1007/s11217-018-9644-1.

Arksey, Hilary, and Lisa O'Malley. 2005. "Scoping Studies: Towards a Methodological Framework." *International Journal of Social Research Methodology* 8(1): 19–32. https://doi.org/10.1080/1364557032000119616.

Atewologun, Doyin, Tinu Cornish, and Fatima Tresh. 2018. "Unconscious Bias Training: An Assessment of the Evidence for Effectiveness." https://www.equalityhumanrights.com/en/publication-download/unconscious-bias-training-assessment-evidence-effectiveness.

Banaji, Mahzarin R., and Anthony G. Greenwald. 1995. "Implicit Gender Stereotyping in Judgments of Fame." *Journal of Personality and Social Psychology* 68(2): 181–98. https://doi.org/10.1037/0022-3514.68.2.181.

Banaji, Mahzarin R., and Anthony G. Greenwald. 2016. *Blindspot: Hidden Biases of Good People*. New York: Random House.

Bandura, Albert. 1988. "Self-Efficacy Conception of Anxiety." *Anxiety Research* 1(2): 77–98. https://doi.org/10.1080/10615808808248222.

Beetz, Johannes, and Veit Schwab. 2017. "Materialist Discourse Analysis: Three Moments and Some Criteria." In *Material Discourse—Materialist Analysis: Approaches in Discourse Studies*, edited by Benno Herzog, J.-Jacques Courtine, Helio Oliveira, Ligia M. B. M. de Araújo, Marco A. A. Ruiz, Roberto L. Baronas, Benjamin Glasson, Laura Pantzerhielm, Manuel Iretzberger and Giorgio Borrelli, 29–43. Lexington Books.

Bhopal, Kalwant. 2020. "Confronting White Privilege: The Importance of Intersectionality in the Sociology of Education." *British Journal of Sociology of Education* 41(6): 807–16. https://doi.org/10.1080/01425692.2020.1755224.

Bhopal, Kalwant, and Clare Pitkin. 2020. "'Same Old Story, Just a Different Policy': Race and Policy Making in Higher Education in the UK." *Race, Ethnicity and Education* 23(4): 530–47. https://doi.org/10.1080/13613324.2020.1718082.

Bickel, Janet, Diane Wara, Barbara F. Atkinson, Lawrence S. Cohen, Michael Dunn, Sharon Hostler, Timothy R. B. Johnson, et al. 2002. "Increasing Women's Leadership in Academic Medicine: Report of the AAMC Project Implementation Committee." *Academic Medicine* 77(10): 1043–61. https://doi.org/10.1097/00001888-200210000-00023.

Brady, Michelle. 2016. "Openings." In *Governing Practices: Neoliberalism, Governmentality, and the Ethnographic Imaginary*, edited by Michelle Brady and Randy K. Lippert, 3–31. University of Toronto Press.

Brewis, Deborah N. 2019. "Duality and Fallibility in Practices of the Self: The 'Inclusive Subject' in Diversity Training." *Organization Studies* 40(1): 93–114. https://doi.org/10.1177/0170840618765554.

Brown, Wendy. (2009). *Edgework: Critical Essays on Knowledge and Politics*. Princeton, NJ: Princeton University Press.

Byars-Winston, Angela, Belinda Gutierrez, Sharon Topp, and Molly Carnes. 2011. "Integrating Theory and Practice to Increase Scientific Workforce Diversity: A Framework for Career Development in Graduate Research Training." *CBE-Life Sciences Education* 10(4): 357–67. https://doi.org/10.1187/cbe.10-12-0145.

Cahn, Peter S. 2017. "Recognizing and Reckoning with Unconscious Bias: A Workshop for Health Professions Faculty Search Committees." *MedEdPORTAL* 13: 10544. https://doi.org/10.15766/mep_2374-8265.10544.

Calanchini, Jimmy, Calvin K. Lai, and Karl Christoph Klauer. 2020. "Reducing Implicit Racial Preferences: III. A Process-Level Examination of Changes in Implicit Preferences." *Journal of Personality and Social Psychology* 121(4): 796–818. https://doi.org/10.1037/pspi0000339.

Calder-Dawe, Octavia, Margaret Wetherell, Maree Martinussen, and Alex Tant. 2021. "Looking on the Bright Side: Positivity Discourse, Affective Practices and New Femininities." *Feminism & Psychology* 31(4): 550–70. https://doi.org/10.1177/09593535211030756.

Carlone, Heidi B., and Angela Johnson. 2007. "Understanding the Science Experiences of Successful Women of Color: Science Identity as an Analytic Lens." *Journal of Research in Science Teaching* 44(8): 1187–218. https://doi.org/10.1002/tea.20237.

Carnes, Molly, Patricia G. Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E. Ford, Patrick Forscher, et al. 2015. "The Effect of an Intervention to Break the Gender Bias Habit for Faculty at One Institution: A Cluster Randomized, Controlled Trial." *Academic Medicine* 90(2): 221–30. https://doi.org/10.1097/ACM.0000000000000552.

Carnes, Molly, Patricia G. Devine, Carol Isaac, Linda Baier Manwell, Cecelia E. Ford, Angela Byars-Winston, Eve Fine, and Jennifer Sheridan. 2012. "Promoting Institutional Change through Bias Literacy." *Journal of Diversity in Higher Education* 5(2): 63–77. https://doi.org/10.1037/a0028128.

Casad, Bettina J., Abdiel J. Flores, and Jessica D. Didway. 2012. "Using the Implicit Association Test as an Unconsciousness Raising Tool in Psychology." *Teaching of Psychology* 40(2): 118–23.

Casad, Bettina J., Jillian E. Franks, Christina E. Garasky, Melinda M. Kittleman, Alanna C. Roesler, Deidre Y. Hall, and Zachary W. Petzel. 2021. "Gender Inequality in Academia: Problems and Solutions for Women Faculty in STEM." *Journal of Neuroscience Research* 99(1): 13–23. https://doi.org/10.1002/jnr.24631.

CASE. 2014. "Improving Diversity in STEM: A Report by the Campaign for Science and Engineering (CaSE)." https://www.sciencecentres.org.uk/resources/promoting-diversity-and-inclusion-stem-engagement/case-improving-diversity-stem/.

Cech, Erin A., and Mary Blair-Loy. 2019. "The Changing Career Trajectories of New Parents in STEM." *Proceedings of the National Academy of Sciences* 116(10): 4182–7. https://doi.org/10.1073/pnas.1810862116.

Chandler, David, and Julian Reid. 2016. *The Neoliberal Subject: Resilience, Adaptation and Vulnerability*. Rowman & Littlefield.

Chang, Shine, Page S. Morahan, Diane Magrane, Deborah Helitzer, Hwa Young Lee, Sharon Newbill, Ho-Lan Peng, Michele Guindani, and Gina Cardinali. 2016. "Retaining Faculty in Academic Medicine: The Impact of Career Development Programs for Women." *Journal of Women's Health* 25(7): 687–96. https://doi.org/10.1089/jwh.2015.5608.

Dean, Mitchell. 2010. *Governmentality: Power and Rule in Modern Society*. 2nd ed. Los Angles, CA, London: Sage.

Devine, Patricia G., Patrick S. Forscher, Anthony J. Austin, and William T. L. Cox. 2012. "Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention." *Journal of Experimental Social Psychology* 48(6): 1267–78. https://doi.org/10.1016/j.jesp.2012.06.003.

Devine, Patricia G., Patrick S. Forscher, William T. L. Cox, Anna Kaatz, Jennifer Sheridan, and Molly Carnes. 2017. "A Gender Bias Habit-Breaking Intervention Led to Increased Hiring of Female Faculty in STEMM Departments." *Journal of Experimental Social Psychology* 73: 211–5. https://doi.org/10.1016/j.jesp.2017.07.002.

Doharty, Nadena, Manuel Madriaga, and Remi Joseph-Salisbury. 2021. "The University Went to 'Decolonise' and All They Brought Back was Lousy Diversity Double-Speak! Critical Race Counter-Stories from Faculty of Colour in 'Decolonial' Times." *Educational Philosophy and Theory* 53(3): 233–44. https://doi.org/10.1080/00131857.2020.1769601.

Dupree, Cydney H., and C. Malik Boykin. 2021. "Racial Inequality in Academia: Systemic Origins, Modern Challenges, and Policy Recommendations." *Policy Insights from the Behavioral and Brain Sciences* 8(1): 11–8. https://doi.org/10.1177/2372732220984183.

Equality Challenge Unit. 2013. "Unconscious Bias and Higher Education." https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/ecu/unconscious-bias-and-higher-education_1579011683.pdf.

Fagan, Colette, and Nina Teasdale. 2021. "Women Professors across STEMM and Non-STEMM Disciplines: Navigating Gendered Spaces and Playing the Academic Game." *Work, Employment & Society* 35(4): 774–92. https://doi.org/10.1177/0950017020916182.

Farkas, Amy H., Eliana Bonifacino, Rose Turner, Sarah A. Tilstra, and Jennifer A. Corbelli. 2019. "Mentorship of Women in Academic Medicine: A Systematic Review." *Journal of General Internal Medicine* 34(7): 1322–9. https://doi.org/10.1007/s11606-019-04955-2.

FitzGerald, Chloë, Angela Martin, Delphine Berner, and Samia Hurst. 2019. "Interventions Designed to Reduce Implicit Prejudices and Implicit Stereotypes in Real World Contexts: A Systematic Review." *BMC Psychology* 7(1): 29. https://doi.org/10.1186/s40359-019-0299-7.

Foucault, Michel. 1991. Remarks on Marx: Conversations with Duccio Trombadori. Semiotext(e) Foreign Agents Series.

Foucault, Michel. 2002. "The Subject and Power." In *Power*, edited by Michel Foucault and James D. Faubion. Essential Works of Foucault, 1954–1984 v. 3. London: Penguin.

Frambach, Janneke M., and Maria Athina Tina Martimianakis. 2017. "The Discomfort of an Educator's Critical Conscience: The Case of Problem-Based Learning and Other Global Industries in Medical Education." *Perspectives on Medical Education* 6(1): 1–4. https://doi.org/10.1007/s40037-016-0325-x.

Gatewood, Elizabeth, Cindy Broholm, Jenna Herman, and Charles Yingling. 2019. "Making the Invisible Visible: Implementing an Implicit Bias Activity in Nursing Education." *Journal of Professional Nursing* 35(6): 447–51. https://doi.org/10.1016/j.profnurs.2019.03.004.

Ghoshal, Raj Andrew, Cameron Lippard, Vanesa Ribas, and Ken Muir. 2013. "Beyond Bigotry: Teaching about Unconscious Prejudice." *Teaching Sociology* 41(2): 130–43. https://doi.org/10.1177/0092055X12446757.

Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102(1): 4.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6): 1464–80. https://doi.org/10.1037/0022-3514.74.6.1464.

Gutiérrez y Muhs, Gabriella, Yolanda Flores Niemann, Carmen G. González, and Angela P. Harris. 2012. Presumed Incompetent: The Intersections of Race and Class for Women in Academia.

Guyan, Kevin, and F. D. Oloyede. 2019. "Equality, Diversity and Inclusion in Research and Innovation: UK Review." https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-EDI-EvidenceReviewUK.pdf.

Hannah, Seth Donal, and Elizabeth Carpenter-song. 2013. "Patrolling Your Blind Spots: Introspection and Public Catharsis in a Medical School Faculty Development Course to Reduce Unconscious Bias in Medicine." *Culture, Medicine and Psychiatry* 37(2): 314–39. https://doi.org/10.1007/s11013-013-9320-4.

Hemmings, Clare. 2012. "Affective Solidarity: Feminist Reflexivity and Political Transformation." *Feminist Theory* 13(2): 147–61. https://doi.org/10.1177/1464700112442643.

Hennes, Erin P., Evava S. Pietri, Corinne A. Moss-Racusin, Katherine A. Mason, John F. Dovidio, Victoria L. Brescoll, April H Bailey, and Jo Handelsman. 2018. "Increasing the Perceived Malleability of Gender Bias Using a Modified Video Intervention for Diversity in STEM (VIDS)." *Group Processes & Intergroup Relations* 21(5): 788–809. https://doi.org/10.1177/1368430218755923.

Herman, Clem, Suzan Lewis, and Anne Laure Humbert. 2013. "Women Scientists and Engineers in European Companies: Putting Motherhood under the Microscope." *Gender, Work and Organization* 20(5): 467–78. https://doi.org/10.1111/j.1468-0432.2012.00596.x.

Hochschild, Arlie Russell. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85(3): 551–75. https://doi.org/10.1086/227049.

Hochschild, Arlie Russell. 2015. "The Managed Heart." In *Working in America*, 47–54. Routledge.

Hunter, Shona. 2015. *Power, Politics and the Emotions: Impossible Governance?* Social Justice. New York, NY: Routledge.

Huston, Wilhelmina M., Charles G. Cranfield, Shari L. Forbes, and Andy Leigh. 2019. "A Sponsorship Action Plan for Increasing Diversity in STEMM." *Ecology and Evolution* 9(5): 2340–5. https://doi.org/10.1002/ece3.4962.

Hutchins, Darvelle, and Marlo Goldstein Hode. 2019. "Exploring Faculty and Staff Development of Cultural Competence through Communicative Learning in an Online Diversity Course." *Journal of Diversity in Higher Education* 14(4): 468–79. https://doi.org/10.1037/dhe0000162.

Jackson, Jessi Lee. 2018. "The Non-Performativity of Implicit Bias Training." *Radical Teacher* 112(112): 46–54. https://doi.org/10.5195/rt.2018.497.

Jackson, Sarah M., Amy L. Hillard, and Tamera R. Schneider. 2014. "Using Implicit Bias Training to Improve Attitudes toward Women in STEM." *Social Psychology of Education* 17(3): 419–38. https://doi.org/10.1007/s11218-014-9259-5.

Jäger, Siegfried. 2001. "Discourse and Knowledge: Theoretical and Methodological Aspects of a Critical Discourse and Dispositive Analysis." *Methods of Critical Discourse Analysis* 1: 32–63.

Johnston, Rowena, Suteeraporn Pinyakorn, and Jintanat Ananworanich. 2016. "Is There Gender Bias in HIV Cure Research? A Case Study of Female Representation at the 2015 HIV Persistence Workshop." *Journal of Virus Eradication* 2(2): 117–20. https://doi.org/10.1016/s2055-6640(20)30477-5.

Kahn, Jonathan. 2018. *Race on the Brain: What Implicit Bias Gets Wrong about the Struggle for Racial Justice*. New York: Columbia University Press.

Kahn, Shulamit, and Donna Ginther. 2017. "Women and STEM: NBER Working Paper Series." https://www.nber.org/system/files/working_papers/w23525/w23525.pdf.

Kelly, Michael P., and Mary Barker. 2016. "Why is Changing Health-Related Behaviour So Difficult?" *Public Health* 136: 109–16. https://doi.org/10.1016/j.puhe.2016.03.030.

Klein, Elise. 2017. "Developing Minds: Psychology, Neoliberalism and Power." In *Concepts for Critical Psychology*. 1 Edition. London, New York: Routledge Taylor & Francis Group.

Kristeva, Julia. 1984. *Powers of Horror: An Essay of Abjection. European Perspectives*. Chichester, New York: Columbia University Press.

Krutkowski, Sebastian, Sarah Taylor-Harman, and Kat Gupta. 2019. "De-Biasing on University Campuses in the Age of Misinformation." *RSR* 48(1): 1208–128. https://doi.org/10.1108/RSR-10-2019-0075.

Lai, Calvin K., Maddalena Marini, Steven A. Lehr, Carlo Cerruti, J.-Elizabeth L. Shin, Jennifer A. Joy-Gaba, Arnold K. Ho, et al. 2014. "Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions." *Journal of Experimental Psychology: General* 143(4): 1765–85. https://doi.org/10.1037/a0036260.

Laver, Kate E., Ivanka J. Prichard, Monica Cations, Ivana Osenk, Kay Govin, and John D. Coveney. 2018. "A Systematic Review of Interventions to Support the Careers of Women in Academic Medicine and Other Disciplines." *BMJ Open* 8(3): e020380. https://doi.org/10.1136/bmjopen-2017-020380.

Lemke, Thomas. 2011. *Foucault, Governmentality, and Critique*. Cultural Politics & the Promise of Democracy. Boulder, CO: Paradigm: Slough: Compass DSA [distributor].

Lemke, Thomas. 2019. *A Critique of Political Reason: Foucault's Analysis of Modern Governmentality*. London, Brooklyn, NY: Verso.

Lewis, Vivian, Camille A. Martina, Michael P. McDermott, Paula M. Trief, Steven R. Goodman, Gene D. Morse, Jennifer G. LaGuardia, Daryl Sharp, and Richard M. Ryan. 2016. "A Randomized Controlled Trial of Mentoring Interventions for Underrepresented Minorities." *Academic Medicine* 91(7): 994–1001. https://doi.org/10.1097/acm.0000000000001056.

Liu, Helena. 2017. "Redeeming Difference in CMS through Anti-racist Feminisms." In *Feminists and Queer Theorists Debate the Future of Critical Management Studies*. Emerald Publishing Limited.

Lorenzini, Daniele. 2018. "Governmentality, Subjectivity, and the Neoliberal Form of Life." *Journal for Cultural Research* 22(2): 154–66. https://doi.org/10.1080/14797585.2018.1461357.

Lueke, Adam, and Bryan Gibson. 2016. "Brief Mindfulness Meditation Reduces Discrimination." *Psychology of Consciousness: Theory, Research, and Practice* 3(1): 34–44. https://doi.org/10.1037/cns0000081.

Magnusson, Eva, and Jeanne Marecek. 2017. "Feminisms, Psychologies, and the Study of Social Life." In *The Palgrave Handbook of Critical Social Psychology*, edited by Brendan Gough. London, UK: Palgrave Macmillan.

Marshall, Alison, Priyanka Sista, Katie Colton, Abra Fant, Howard Kim, Patrick Lank, and Danielle McCarthy. 2020. "Women's Night in Emergency Medicine Mentorship Program: A SWOT Analysis." *Western Journal of Emergency Medicine* 21(1): 37–41. https://doi.org/10.5811/westjem.2019.11.44433.

Martin, Graham P., and Justin Waring. 2018. "Realising Governmentality: Pastoral Power, Governmental Discourse and the (Re) Constitution of Subjectivities." *The Sociological Review* 66(6): 1292–308. https://doi.org/10.1177/0038026118755616.

Mavin, Sharon, and Jannine Williams. 2013. "Women's Impact on Women's Careers in Management: Queen Bees, Female Misogyny, Negative Intra-Relations and Solidarity Behaviours." In *Handbook of Research on Promoting Women's Careers*. Edward Elgar Publishing.

Mays, Nicholas, Emilie Roberts, and Jennie Popay. 2004. "Synthesising Research Evidence." In *Studying the Organization and Delivery of Health Services*, 200–32. Routledge.

McGee, Ebony O., and Lydia Bentley. 2017. "The Troubled Success of Black Women in STEM." *Cognition and Instruction* 35(4): 265–89. https://doi.org/10.1080/07370008.2017.1355211.

McMahon, John. 2015. "Behavioral Economics as Neoliberalism: Producing and Governing Homo Economicus." *Contemporary Political Theory* 14(2): 137–58. https://doi.org/10.1057/cpt.2014.14.

Millner, Naomi. 2022. "Unsettling Feelings in the Classroom: Scaffolding Pedagogies of Discomfort as Part of Decolonising Human Geography in Higher Education." *Journal of Geography in Higher Education*: 1–20. https://doi.org/10.1080/03098265.2021.2004391.

Moonesirust, Elham, and Andrew D. Brown. 2021. "Company Towns and the Governmentality of Desired Identities." *Human Relations* 74(4): 502–26. https://doi.org/10.1177/0018726719887220.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109(41): 16474–9. https://doi.org/10.1073/pnas.1211286109.

Moss-Racusin, Corinne A., Evava S. Pietri, Erin P. Hennes, John F. Dovidio, Victoria L. Brescoll, Gina Roussos, and Jo Handelsman. 2018. "Reducing STEM Gender Bias with VIDS (Video Interventions for Diversity in STEM)." *Journal of Experimental Psychology: Applied* 24(2): 236–60. https://doi.org/10.1037/xap0000144.

Nelson, Laura K., and Kathrin Zippel. 2021. "From Theory to Practice and Back: How the Concept of Implicit Bias was Implemented in Academe, and What this Means for Gender Theories of Organizational Change." *Gender & Society* 35(3): 330–57. https://doi.org/10.1177/08912432211000335.

Noon, Mike. 2018. "Pointless Diversity Training: Unconscious Bias, New Racism and Agency." *Work, Employment & Society* 32(1): 198–209. https://doi.org/10.1177/0950017017719841.

Ong, Maria, Janet M. Smith, and Lily T. Ko. 2018. "Counterspaces for Women of Color in STEM Higher Education: Marginal and Central Spaces for Persistence and Success." *Journal of Research in Science Teaching* 55(2): 206–45. https://doi.org/10.1002/tea.21417.

Oswick, Cliff, and Mike Noon. 2014. "Discourses of Diversity, Equality and Inclusion: Trenchant Formulations or Transient Fashions?" *British Journal of Management* 25(1): 23–39. https://doi.org/10.1111/j.1467-8551.2012.00830.x.

Parker, Ian. 2020a. "Psychology through Critical Auto-Ethnography: Academic Discipline, Professional Practice and Reflexive History by Ian Parker." *Awry: Journal of Critical Psychology* 1(1): 101–5. https://awryjcp.com/index.php/awry/article/view/16.

Parker, Ian. 2020b. *Psychology through Critical Auto-Ethnography: Academic Discipline, Professional Practice and Reflexive History*. 1st edition. London: Routledge.

Prochaska, James O., and Wayne F. Velicer. 1997. "The Transtheoretical Model of Health Behavior Change." *American Journal of Health Promotion* 12(1): 38–48. https://doi.org/10.4278/0890-1171-12.1.38.

Project Implicit. 2021. "Take a Test.". https://implicit.harvard.edu/implicit/takeatest.html.

Pullen, Alison, Sheena Vachhani, Suzanne Gagnon, and Nelarine Cornelius. 2017. "Critical Diversity, Philosophy and Praxis."

Rees, Teresa. 2001. "Mainstreaming Gender Equality in Science in the European Union: The 'ETAN Report." *Gender and Education* 13(3): 243–60. https://doi.org/10.1080/09540250120063544.

Rhoton, Laura A. 2011. "Distancing as a Gendered Barrier: Understanding Women Scientists' Gender Practices." *Gender & Society* 25(6): 696–716. https://doi.org/10.1177/0891243211422717.

Riley, Sarah, Adrienne Evans, Emma Anderson, and Martine Robson. 2019. "The Gendered Nature of Self-Help." *Feminism & Psychology* 29(1): 3–18. https://doi.org/10.1177/0959353519826162.

Riley, Sarah C. E. 2002. "Constructions of Equality and Discrimination in Professional Men's Talk." *British Journal of Social Psychology* 41(3): 443–61. https://doi.org/10.1348/014466602760344304.

Rodriguez, Jenny K., Evangelina Holvino, Joyce K. Fletcher, and Stella M. Nkomo. 2016. "The Theory and Praxis of Intersectionality in Work and Organizations: Where Do We Go from Here?" *Gender, Work and Organization* 23(3): 201–22. https://doi.org/10.1111/gwao.12131.

Rose, Nikolas. 1998. *Inventing Our Selves: Psychology, Power, and Personhood*. Cambridge University Press.

Rose, Nikolas. 2001. "The Politics of Life Itself." *Theory, Culture & Society* 18(6): 1–30. https://doi.org/10.1177/02632760122052020.

Rose, Nikolas S. 1985. *The Psychological Complex: Psychology, Politics, and Society in England, 1869–1939*. Routledge & Kegan Paul.

Rottenberg, Catherine. 2014. "The Rise of Neoliberal Feminism." *Cultural Studies* 28(3): 418–37. https://doi.org/10.1080/09502386.2013.857361.

Rubini, Monica, and Michela Menegatti. 2014. "Hindering Women's Careers in Academia: Gender Linguistic Bias in Personnel Selection." *Journal of Language and Social Psychology* 33(6): 632–50. https://doi.org/10.1177/0261927x14542436.

Saetermoe, Carrie L., Gabriela Chavira, Crist S. Khachikian, David Boyns, and Beverly Cabello. 2017. "Critical Race Theory as a Bridge in Science Training: The California State University, Northridge BUILD PODER Program." *BMC Proceedings* 11(Suppl 12): 21. https://doi.org/10.1186/s12919-017-0089-2.

Sevo, Ruta, and Daryl E. Chubin. 2010. "Bias Literacy: A Review of Concepts in Research on Gender Discrimination and the US Context." In *Women in Engineering, Science and Technology: Education and Career Challenges*, 21–54.

Sian, Katy P. 2019. *Navigating Institutional Racism in British Universities*. Mapping Global Racisms. Cham, Switzerland: Palgrave Macmillan.

Silver, Julie K., Cheri A. Blauwet, Saurabha Bhatnagar, Chloe S. Slocum, Adam S. Tenforde, Jeffrey C. Schneider, Ross D. Zafonte, et al. 2018. "Women Physicians Are Underrepresented in Recognition Awards from the Association of Academic Physiatrists." *American Journal of Physical Medicine & Rehabilitation* 97(1): 34–40. https://doi.org/10.1097/phm.0000000000000792.

Souto-Manning, Mariana, and Nichole Ray. 2007. "Beyond Survival in the Ivory Tower: Black and Brown Women's Living Narratives." *Equity & Excellence in Education* 40(4): 280–90. https://doi.org/10.1080/10665680701588174.

Stone, Jeff, Gordon B. Moskowitz, Colin A. Zestcott, and Katherine J. Wolsiefer. 2020. "Testing Active Learning Workshops for Reducing Implicit Stereotyping of Hispanics by Majority and Minority Group Medical Students." *Stigma and Health* 5(1): 94–103. https://doi.org/10.1037/sah0000179.

Tate, Shirley Anne, and Damien Page. 2018. "Whiteliness and Institutional Racism: Hiding Behind (Un)Conscious Bias." *Ethics and Education* 13(1): 141–55. https://doi.org/10.1080/17449642.2018.1428718.

Thébaud, Sarah, and Catherine J. Taylor. 2021. "The Specter of Motherhood: Culture and the Production of Gendered Career Aspirations in Science and Engineering." *Gender & Society* 35(3): 395–421. https://doi.org/10.1177/08912432211006037.

The Royal Society. 2021. "All-Party Parliamentary Group on Diversity and Inclusion in STEM Inquiry into Equity in the STEM Workforce." https://royalsociety.org/-/media/policy/Publications/2021/02-01-21-royal-society-submission-appg-diversity-and-inclusion-in-STEM.pdf.

Townley, Barbara. 1997. "The Institutional Logic of Performance Appraisal." *Organization Studies* 18(2): 261–85. https://doi.org/10.1177/017084069701800204.

Triantafillou, Peter. 2012. *New Forms of Governing: A Foucauldian Inspired Analysis*. Basingstoke: Palgrave Macmillan.

Tzanakou, Charikleia, and Ruth Pearce. 2019. "Moderate Feminism Within or against the Neoliberal University? The Example of Athena SWAN." *Gender, Work and Organization* 26(8): 1191–211. https://doi.org/10.1111/gwao.12336.

UK Cabinet Office. 2020. "Written Ministerial Statement on Unconscious Bias Training." News Release. https://www.gov.uk/government/news/written-ministerial-statement-on-unconscious-bias-training.

UKRI. 2021. "UKRI Equality Diversity and Inclusion Strategy: Draft for Consultation." https://www.ukri.org/publications/equality-diversity-and-inclusion-strategy-draft-for-consultation/ukri-equality-diversity-and-inclusion-strategy-draft-for-consultation/.

University of Nottingham. 2022. "Research Excellence Framework (REF) 2021 - Unconscious Bias Training for Output Reviewers." https://training.nottingham.ac.uk/Course?courseref=eLREF2021UB.

van den Brink, Marieke, and Lineke Stobbe. 2014. "The Support Paradox: Overcoming Dilemmas in Gender Equality Programs." *Scandinavian Journal of Management* 30(2): 163–74. https://doi.org/10.1016/j.scaman.2013.07.001.

Vos, Jan de. 2012. *Psychologization in Times of Globalisation*. Hove, New York, NY: Routledge.

West, Tahira J., Kimberly Loomer, and Tasha R. Wyatt. 2019. "How Diverse is Your Universe? An Activity for Students to Reflect on Ethnoracial Diversity during Orientation." *MedEdPORTAL* 15: 10840. https://doi.org/10.15766/mep_2374-8265.10840.

Wetherell, Margaret. 2012. *Affect and Emotion: A New Social Science Understanding*. London: Sage.

White-Davis, Tanya, Jennifer Edgoose, Joedrecka S. Brown Speights, Kathryn Fraser, Jeffrey M. Ring, Jessica Guh, and George W. Saba. 2018. "Addressing Racism in Medical Education: An Interactive Training Module." *Family Medicine* 50(5): 364–8. https://doi.org/10.22454/FamMed.2018.875510.

Wigginton, Britta, and Michelle N. Lafrance. 2019. "Learning Critical Feminist Research: A Brief Introduction to Feminist Epistemologies and Methodologies." *Feminism & Psychology* 36(3): 095935351986605. https://doi.org/10.1177/0959353519866058.

Willig, Carla. 2014. *The SAGE Handbook of Qualitative Data Analysis*. London: Sage Publications Ltd.

Xiao, Yunyu, Edward Pinkney, Terry Kit Fong Au, and Paul Siu Fai Yip. 2020. "Athena Swan and Gender Diversity: A UK-Based Retrospective Cohort Study." *BMJ Open* 10(2): e032915. https://doi.org/10.1136/bmjopen-2019-032915.

AUTHOR BIOGRAPHIES

Christian Möller is a research associate in the School of Social Work and Social Policy at the University of Strathclyde and associate lecturer at The Open University, UK. He is a critical psychologist with research interests in critical discourse analysis, visual methods, and the links between psychology and neoliberalism.

Saffron Passam is a lecturer in Psychology at Aberystwyth University, UK. Her research interests include social interactions and the workplace, especially issues relating to gender, race, disability, and class. Recently, Saffron contributed to a successful Engineering and Physical Sciences Research Council (EPSRC) bid that has investigated matters of unconscious bias and inclusion in academia. She is an active member of the Center for Critical Psychology at Aberystwyth and her work challenges notions of meritocracy and makes visible the potential for troubled and unequal practices in organizational systems which otherwise hold a commitment to equality.

Sarah Riley is Professor in Critical Health Psychology at Massey University, Aotearoa New Zealand, and the director of its Health Psychology master's program. Her research examines discourse, affect and materiality in relation to digital technology, subjectivity, gender, bodies, and neoliberalism.

Martine Robson is a lecturer in Psychology at Aberystwyth University, UK. Her work focuses on how people negotiate individualistic health-related lifestyle advice and uses poststructuralist theory to examine the ways in which people adopt and resist neoliberal healthism. Martine's research also encompasses inclusivity in the workplace in an EPSRC project exploring inequality in academic workplaces. She is the director of the Center for Critical Psychology at Aberystwyth and works with clinicians in the NHS to apply poststructuralist concepts in clinical settings.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Möller, Christian, Saffron Passam, Sarah Riley, and Martine Robson. 2023. "All Inside Our Heads? A Critical Discursive Review of Unconscious Bias Training in the Sciences." *Gender, Work & Organization*: 1–24. https://doi.org/10.1111/gwao.13028.