Intervention Methods for Implicit Social Cognition

Authors: Guo Xiaoli, Chang Junyao, Shama Ajie, Yang Ziyan, Yang Ziyan

Date: 2025-05-10T16:30:51+00:00

Abstract

Since the concept of implicit social cognition was proposed in 1995, this field has attracted extensive research, covering various aspects including implicit attitudes, implicit stereotypes, and implicit self-esteem. Early studies primarily focused on the measurement of implicit social cognition and its impact on individuals. Research has demonstrated that negative implicit social cognition (such as implicit bias, implicit stereotypes, and low implicit self-esteem) exerts a series of adverse effects on people' s psychology and behavior. In recent years, the research focus has gradually shifted toward effective interventions for negative implicit social cognition, aiming to mitigate its detrimental consequences. Currently, nine commonly used effective intervention strategies include evaluative conditioning, approach-avoidance training, intergroup contact, exposure to counter-stereotypical exemplars, implicit bias training, implementation intentions, targeted memory reactivation, emotion induction, and mindfulness meditation. This paper provides a systematic overview of these intervention methods and their effectiveness, conducting a meta-analysis of the effect sizes for each method. The results indicate that these methods exhibit small-to-medium effect sizes, with approach-avoidance training, implicit bias training, and emotion induction interventions being relatively more effective. The aforementioned intervention methods demonstrate certain short-term effectiveness but lack long-term durability. Future research could employ combined multi-method approaches to enhance intervention effectiveness and durability. Additionally, future implicit social cognition interventions could become more efficient and convenient with the support of artificial intelligence technology.

Full Text

Preamble

Interventions for Implicit Social Cognition

Guo Xiaoli[1],[2], Chang Junyao[1],[2], Sha Maajie[1],[2], Yang Ziyan[1],[2]
(1 State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)
(2 Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Since the concept of implicit social cognition was introduced in 1995, this field has attracted extensive research covering implicit attitudes, implicit stereotypes, and implicit self-esteem. Early studies primarily focused on measuring implicit social cognition and its effects on individuals. Research has shown that negative implicit social cognition (such as implicit bias, implicit stereotypes, and low implicit self-esteem) exerts a range of detrimental effects on people's psychology and behavior. In recent years, research attention has gradually shifted toward effectively intervening in negative implicit social cognition to reduce its adverse consequences. Currently, nine commonly used and effective intervention methods include evaluative conditioning, approach-avoidance training, intergroup contact, exposure to counter-stereotypical exemplars, unconscious bias training, implementation intentions, targeted memory reactivation, emotion induction, and mindfulness meditation. This article provides a systematic review of these intervention methods and their effectiveness, with meta-analyses of effect sizes for each method. Results indicate that these methods demonstrate small to medium effect sizes, with approach-avoidance training, unconscious bias training, and emotion induction interventions showing relatively greater effectiveness. While these interventions demonstrate short-term efficacy, they lack long-term durability. Future research could combine multiple methods to enhance both intervention effectiveness and durability. Additionally, future implicit social cognition interventions could become more efficient and convenient with support from artificial intelligence technologies.

Keywords: implicit social cognition, Implicit Association Test, implicit bias, implicit stereotypes, intervention strategies

With the development of social cognition research, people have gradually realized that many psychological processes and behaviors are not entirely driven by explicit consciousness but are deeply rooted in individuals' unconscious levels (Nisbett & Wilson, 1977). In the 1990s, the concept of "implicit social cognition" provided a new perspective for understanding the underlying psychological mechanisms of human behavior. According to implicit social cognition theory, individuals' social psychology and behavior are often influenced by unconscious, automatic psychological processes (Freud, 1960). Implicit social cognition has gradually become a focal point in social psychology, encompassing implicit attitudes, implicit stereotypes, and implicit self-esteem. Numerous studies have revealed that implicit social cognition plays an important role in individual psychology and behavior, and even has broad impacts on macro-level social systems including education, employment, and justice (Greenwald et al., 2022; Greenwald & Lai, 2020; Yang et al., 2015).

1 Received: March 16, 2025

Funding: National Natural Science Foundation of China General Program, 32371130
Corresponding Author: Yang Ziyan, E-mail: yangzy@psych.ac.cn

As understanding of implicit social cognition deepens, particularly regarding the potential negative consequences of negative implicit social cognition (such as implicit bias, implicit stereotypes, and low implicit self-esteem)—including prejudice, discrimination, and social injustice (Greenwald & Lai, 2020; Yang et al., 2015)—researchers have increasingly focused on how to effectively intervene in these implicit social cognitions to reduce their adverse effects. To this end, researchers have developed a series of intervention methods for implicit social cognition, including evaluative conditioning, approach-avoidance training, intergroup contact, exposure to counter-stereotypical exemplars, and others, with empirical studies validating the effectiveness of these methods. This article systematically reviews effective intervention methods for implicit social cognition and reveals the intervention effects of various approaches through meta-analysis. By integrating existing intervention methods, this paper aims to provide a foundation for future research and practice in implicit social cognition intervention and to explore possible directions for improving intervention effectiveness.

2 The Connotation and Measurement of Implicit Social Cognition

Implicit social cognition primarily describes automatic, unconscious psychological processes that influence individuals' psychology and behavior without their awareness (Greenwald & Banaji, 1995). It mainly includes implicit attitudes, implicit stereotypes, and implicit self-esteem. Implicit attitudes refer to individuals' unconscious, automatic evaluative reactions or emotional tendencies that typically influence specific objects, groups, or concepts (Greenwald et al., 2002; Wilson et al., 2000). Implicit bias is often regarded as a special manifestation of implicit attitudes or negative implicit attitudes (Gawronski, 2024; Lai et al., 2013). Research indicates that implicit bias is widespread, with individuals often exhibiting biased behavior unconsciously and typically failing to recognize that their behavior is influenced by bias (Gawronski et al., 2022; Greenwald, Dasgupta, et al., 2022). Implicit stereotypes refer to individuals' automatic, generalized cognitions or beliefs about specific groups that exist unconsciously and influence judgments and behaviors toward group members (Greenwald & Banaji, 1995). Implicit stereotypes are prevalent across social categories including gender, race, age, and occupation. Implicit self-esteem represents individuals' automatic, unconscious self-evaluative emotions (Falk & Heine, 2015; Greenwald et al., 2002) and plays an important role in mental health (Yang et al., 2015).

Implicit social cognition is typically measured using indirect methods, with the most common being the Implicit Association Test (IAT) (Greenwald, Brendl, et

al., 2022). The IAT employs a computer-based categorization task, measuring the difference in reaction times between compatible tasks (e.g., "White-positive") and incompatible tasks (e.g., "Black-negative") to assess automatic associations between concepts and attributes, thereby indirectly measuring individuals' implicit attitudes, stereotypes, and self-esteem (Greenwald et al., 1998). As the most popular and effective tool in implicit measurement, the IAT is widely used in implicit social cognition research (Cai & Wu, 2021; Jin et al., 2022; Zhu & Liao, 2020).

3 The Stability and Variability of Implicit Social Cognition

Whether implicit social cognition is stable or variable has been a controversial issue of great concern in social psychology. Early slow-learning process theories suggest that individuals gradually develop implicit biases through incremental learning throughout their lives, implying that implicit social cognition changes with natural development (E. R. Smith & DeCoster, 2000). In contrast, category learning theory proposes that once individuals master a social category concept, all associated connections form automatically and remain stable, suggesting that implicit social cognition should be stable and unchanging during individual development (Baron, 2015; Dunham et al., 2008, 2013).

In recent years, increasing research has found that implicit social cognition is variable at both individual and group levels (Marini et al., 2024; Vuletich & Payne, 2019). Changes in implicit social cognition are influenced by both short-term situational factors and long-term cultural factors. For example, specific social information (priming multiculturalism) can trigger short-term changes in implicit attitudes, while lasting attitude changes require long-term influence from education, policy, and culture (Lai et al., 2013). At the individual level, Cvencek et al. found that implicit stereotypes (e.g., gender-academics) deepen during development, particularly among school-age children and adolescents, where individuals' implicit stereotypes gradually strengthen with age (Cvencek et al., 2014, 2015). Cai et al.' s longitudinal research further revealed that implicit social cognition possesses both stability and potential for change over time. Specifically, implicit self-esteem showed lower stability with an overall declining trend; gender-academics related implicit stereotypes demonstrated strong stability; while implicit racial attitudes showed moderate stability with high heterogeneity in individual changes (Cai et al., 2021). At the group level, Charlesworth and colleagues' series of studies revealed that explicit bias has generally decreased, while the change pattern of implicit bias is more complex, influenced by both individual development and sociocultural environments (Charlesworth & Banaji, 2019, 2022a, 2021). Among these, implicit racial bias is more susceptible to change under long-term cultural and educational interventions, whereas gender bias is more stable (Charlesworth & Banaji, 2022b). Cognitive neuroscience research has revealed the physiological mechanisms underlying changes in implicit social cognition, with brain regions such as the prefrontal cortex and amygdala showing activity closely related to changes in implicit bias (Cai & Wu,

2021; Gao et al., 2023).

In summary, implicit social cognition possesses a certain degree of stability while also demonstrating variability due to situational factors. Given this, researchers have dedicated efforts to exploring various intervention strategies to reduce negative implicit social cognition (such as implicit bias) and enhance positive implicit social cognition (such as implicit self-esteem). The following sections systematically introduce the main current intervention methods for implicit social cognition.

4 Intervention Methods for Implicit Social Cognition and Their Effects

Based on existing research, this article has compiled the main intervention methods for implicit social cognition, with detailed introductions to various methods and their effects below. This article systematically summarizes the intervention content, principles, effectiveness, advantages, and disadvantages of these methods (Table 1). We further conducted meta-analyses of effect sizes for each method, first converting effect sizes reported in original studies to Cohen' s d, then performing meta-analyses based on these Cohen' s d values using the Major module in Jamovi software (Table 1).

Table 1 Summary of Intervention Methods for Implicit Social Cognition

Intervention Method	Theoretical Basis	Intervention Content	Effect Size Range	Meta-analytic Effect Size
Evaluative Conditioning	Classical Conditioning	Implicit racial bias	0.18~1.24	0.52
		Implicit stereotypes	0.36~0.59	
Approach-Avoidance Training	Behaviorist Theory	Implicit racial bias	0.11~2.05	0.58
		Implicit academic attitudes	0.11~1.02	
Intergroup Contact	Intergroup Interaction Theory	Implicit racial bias	0.02~1.17	0.46
		Implicit group bias	0.32~1.34	

Intervention Method	Theoretical Basis	Intervention Content	Effect Size Range	Meta-analytic Effect Size
Exposure to Counter-stereotypical Exemplars	Cognitive Reconstruction Theory	Implicit racial bias	0.03~0.66	0.32
		Implicit age bias	0.46~1.67	
Unconscious Bias Training	Cognitive-Behavioral Theory	Implicit racial bias	0.39~0.56	0.54
		Implicit stereotypes	0.34~1.74	
Implementation Intentions	Goal Setting Theory	Implicit racial bias	0.23~0.44	0.38
		Implicit stereotypes	0.23~0.81	
Targeted Memory Reactivation	Memory Consolidation Theory	Implicit racial bias	0.04~0.85	0.42
		Implicit stereotypes	0.46~0.8	
Emotion Induction	Cognitive Reconstruction + Emotion Regulation	Implicit racial bias	0.54~1.04	0.56
		Implicit stereotypes	0.17~0.72	
Mindfulness Meditation	Cognitive Reconstruction + Mindfulness	Implicit racial bias	0.17~0.51	0.41
		Implicit stereotypes	0.46~0.69	

4.1 Evaluative Conditioning (EC)

Evaluative Conditioning (EC) is a procedure based on Pavlovian classical conditioning theory, widely applied in attitude formation and change. Its core strategy involves repeatedly pairing a conditional stimulus (CS) with an unconditional stimulus (US) that has clear emotional valence, thereby prompting changes in individuals' attitudes or evaluations toward the CS (Moran et al.,

2023). Numerous empirical studies have validated EC' s effectiveness in implicit bias intervention. For example, Olson and colleagues paired Black faces (CS) with positive words (US) and found that participants' implicit attitudes toward Black people became more positive, demonstrating that EC can reduce implicit racial bias (Olson & Fazio, 2006). Similarly, establishing connections between positive emotions and specific groups (such as Black people) can effectively reduce implicit bias toward those groups and enhance positive emotional connections with them (Phills et al., 2019), improve intergroup implicit attitudes, and further promote attitude change (Kurdi et al., 2019). Additional research shows that EC can reduce implicit bias toward specific racial groups, such as bias against Muslim and Arab populations (French et al., 2013). Lai et al. validated EC' s effectiveness in reducing implicit racial bias through a series of studies but found the intervention' s durability limited, with effects not exceeding 24 hours (Calanchini et al., 2021; Lai et al., 2016; Lai et al., 2014).

EC' s application extends beyond implicit bias; it can also effectively reduce implicit stereotypes while decreasing implicit bias (Phills et al., 2020). For instance, "counter-stereotype training" based on EC logic reduces implicit stereotypes by presenting a series of stimulus combinations that contradict stereotypes, such as associating the social category "female" with trait words like "strong" and requiring participants to respond to these non-stereotypical combinations to form counter-stereotypical associations. Results show that counter-stereotype training successfully reduces implicit gender and racial stereotypes (Gawronski et al., 2008).

EC can also enhance specific implicit attitudes. For example, Qiu and Zhang found that EC can improve people' s implicit attitudes toward exercise, though its impact on actual exercise behavior is limited (Qiu & Zhang, 2020; Zhang et al., 2023). Jeon et al. (2019) demonstrated that EC can effectively improve adolescents' implicit attitudes toward violent behavior. Additionally, EC has proven effective in changing individuals' implicit attitudes toward alcohol consumption (Zerhouni et al., 2019), diet (Shaw et al., 2016), and specific brands (Charlesworth et al., 2020). EC can also enhance implicit self-esteem. Multiple studies show that pairing the self-concept "I" with positive trait words can effectively boost individuals' implicit self-esteem (Baccus et al., 2004; Dijksterhuis, 2004; Grumm et al., 2009). Recent research indicates that EC' s enhancement effect on implicit self-esteem is effective not only in normal populations but also in clinical groups (such as patients with paranoia) (Trucharte et al., 2024). In summary, EC can effectively change implicit social cognition by associating positive or negative emotional stimuli with specific target objects, thereby further influencing individual behavior (Luo et al., 2024; Tello et al., 2020).

4.2 Approach-Avoidance Training (AAT)

Approach-Avoidance Training (AAT) is a method that influences individuals' attitudes through physical approach or avoidance of specific stimuli. By training participants to approach Black people (pulling pictures of Black faces toward

themselves using a joystick) and avoid White people (pushing pictures of White faces away), followed by measuring racial bias with a racial IAT, results show that repeated approach toward a social group can improve implicit attitudes toward that group. This training also promotes positive nonverbal behaviors (such as smiling and eye contact) in actual interracial interactions (Kawakami et al., 2007). AAT' s effectiveness in reducing implicit racial bias has been supported by subsequent research (Van Dessel et al., 2016, 2020).

Phills et al. further expanded AAT' s application by examining how the match between strategy (approach or avoidance) and context (positive or negative) influences intervention effectiveness. Results show that approach strategies significantly reduce implicit bias in positive contexts, while avoidance strategies are more effective in negative contexts, highlighting the importance of strategy-context matching (Phills et al., 2011). Adding specific instructions during AAT, known as approach-avoidance instructions (AA instructions), can also enhance intervention effects. AA instructions refer to specific operational directives guiding participants to approach or avoid relevant stimuli during experiments or training to ensure compliance with predetermined rules. Recent research found that compared to traditional AAT, adding AA instructions produces larger effects in changing individuals' implicit evaluations of unfamiliar food brands (C. T. Smith et al., 2020), further expanding AAT' s potential in domain-specific implicit attitude interventions. In summary, approach-avoidance training can effectively intervene in implicit bias and broader implicit attitudes by guiding individuals to approach or avoid specific stimuli, with instructions providing additional benefits at critical moments.

4.3 Intergroup Contact

Intergroup contact theory, as an important strategy for reducing intergroup prejudice, has been extended to implicit bias intervention (Paluck et al., 2019; Pettigrew & Tropp, 2006). Intergroup contact can be divided into direct and indirect contact. Extensive research shows that both direct and indirect contact can effectively reduce intergroup implicit bias (Bond et al., 2023). Regarding direct contact, Shook et al. (2008) conducted a natural experiment in a university dormitory system, randomly assigning White freshmen to live with either African American or White roommates. They found that students living with different-race roommates showed significant improvements in implicit racial attitudes and bias, while those living with same-race roommates showed no significant changes. Similarly, Wang and Zhang (2020) assigned software developers to research and development teams with different gender ratios and found that direct contact between genders could reduce implicit gender bias in software development careers. Other research reveals that contact quality and frequency significantly influence intervention effectiveness (Farmer et al., 2020; Onyeador et al., 2020; Wittlin et al., 2019). A recent longitudinal study showed that high-quality, frequent contact can significantly reduce implicit bias and predict improvements in outgroup implicit attitudes, providing practical guidance

for intergroup harmony (Di Bernardo et al., 2022). Additionally, when implementing intergroup contact interventions, positive interaction methods should be encouraged while minimizing negative contact experiences (Barlow et al., 2012).

Direct contact can effectively intervene in implicit bias. For situations where direct contact is impossible, researchers have developed indirect contact strategies (such as imagined intergroup contact and virtual reality contact). Imagined intergroup contact can effectively reduce implicit bias toward specific groups (Stathi et al., 2014; Turner & Crisp, 2010; West et al., 2017). For example, imagined intergroup contact has been used to reduce implicit bias toward migrant workers, homosexual individuals, and obese populations, though contact type and individuals' prior contact experience affect intervention effectiveness. With the development of Virtual Reality (VR) technology, researchers have attempted to extend intergroup contact to virtual environments. Tassinari et al. studied the impact of intergroup contact in VR on implicit attitudes. Participants immersed themselves in virtual environments through VR, interacting with Black (intergroup contact) or White (intragroup contact) players as White avatars. The study found that cooperative contact with Black individuals through VR could improve implicit attitudes toward Black people (Tassinari et al., 2024). These findings support intergroup contact theory and extend its application to virtual reality environments (Doan et al., 2021; Sullivan-Bissett, 2023), simulating positive intergroup interactions that are difficult to achieve in the real world to reduce implicit bias. In summary, both direct and indirect intergroup contact have been widely applied to reduce intergroup implicit bias, and intervention effects can extend to implicit stereotype domains (Eisenstadt et al., 2024). Specifically, enhancing positive intergroup interactions can increase positive emotions toward immigrants, reduce negative implicit stereotypes about immigrants, and alleviate perceived threats from immigrants (Duque et al., 2024).

4.4 Exposure to Counter-stereotypical Exemplars

Exposure to counter-stereotypical exemplars reduces implicit bias and stereotypes toward specific groups by contacting specific images that contradict stereotypes, including four specific intervention methods. First, **Vivid Counter-stereotypic Scenario** is an intervention method that uses emotional and situational immersion to contact counter-stereotypical scenes. In vivid counter-stereotypical scenarios, participants read counter-stereotypical stories from a first-person perspective (e.g., stories where Black people are positive figures, White people are negative figures, and Black heroes save "me" from attacks by White criminals). This method can reduce implicit racial bias toward Black people (Marini et al., 2011). Second, **Practicing an IAT with Counter-stereotypical Exemplars** combines the IAT with counter-stereotypical exemplars. Lai et al. (2014) used positive typical Black and negative typical White exemplars in experiments to strengthen positive associations with Black people and negative associations with White people, effectively reducing participants'

implicit racial bias toward Black people. Third, **Shifting Group Boundaries Through Competition** aims to change perceptions of different racial groups through simulated teamwork and competition. In a simulated dodgeball game, participants cooperated with Black teammates against White opponents, where Black teammates displayed positive behaviors while White opponents showed negative behaviors. Research shows this method can effectively reduce implicit bias toward Black groups (Lai et al., 2014). Fourth, **Shifting Group Affiliations Under Threat** is a method that intervenes in implicit attitudes through extreme threat scenarios. Participants read a story about a post-nuclear war threat scenario (where Black characters were depicted as "close friends" with positive survival skills, while White characters were "terrible enemies" attempting to destroy the camp). Results showed this approach could reduce participants' negative bias toward Black people (Park et al., 2007). Exposure to counter-stereotypical exemplars is also applicable for reducing implicit racial bias in children, though intervention effects are weaker in younger children (ages 5-8) than in older children (ages 9-12) (Gonzalez et al., 2017, 2020). Other research indicates that relying solely on exposure to counter-stereotypical role models is sometimes insufficient to effectively reduce implicit bias, requiring more personalized training programs to specifically target implicit bias reduction (Qian et al., 2022; Qian et al., 2017).

In summary, counter-stereotypical exemplars show some effectiveness in reducing implicit racial bias. However, existing intervention methods have primarily been validated in experimental settings, with their effectiveness and durability in real-world contexts not yet fully evaluated. Some research also notes these methods lack lasting effects (FitzGerald et al., 2019; Lai et al., 2016), and single exposure to counter-stereotypical exemplars is sometimes ineffective, requiring combination with other methods.

4.5 Unconscious Bias Training (UBT)

Unconscious Bias Training (UBT), also known as Implicit Bias-Oriented Diversity Training, reduces implicit bias by combining multiple cognitive intervention strategies to enhance individuals' awareness and control of implicit bias (Hagiwara et al., 2024; Paluck et al., 2021; Atewologun et al., 2018). UBT includes two core components: education and training. The education component conveys the concept of implicit bias to participants, helping them understand the existence of implicit bias and its impact on personal decision-making and social behavior, particularly potential negative consequences. The training component includes several intervention strategies designed to encourage participants to re-examine outgroups from different perspectives and reduce implicit bias toward specific groups. UBT has broad application and 致力于将干预应用于实际 (Devine & Ash, 2022; Moller et al., 2024). Several commonly used intervention measures in UBT are introduced below.

4.5.1 Situational Attribution Training (SAT) Situational Attribution Training guides people to pay more attention to external situational factors, reducing the tendency to attribute outgroup members' negative behaviors to internal traits—the fundamental attribution error (Pettigrew, 1979). For example, Stewart et al. found that White participants who received SAT were more inclined to use situational factors to explain Black men' s negative behaviors, showing fewer implicit racial stereotypes (T. L. Stewart et al., 2010), and that situational attribution can reduce intergroup implicit bias and subsequent negative behaviors (Levontin et al., 2013). Stewart et al.' s recent research used more intensive training designs, increasing training depth and breadth, and found that SAT' s effects remained significant after 24 hours and could generalize to negative implicit stereotypes not involved in training. Recent studies demonstrate SAT' s potential for durability and suggest that repeated training or combination with other interventions may further extend its effects (T. L. Stewart et al., 2022). Overall, SAT' s durability and broad applicability make it an important method for implicit bias intervention.

4.5.2 Prejudice Habit-Breaking Intervention (PHBI) The Prejudice Habit-Breaking Intervention theory views implicit bias and stereotypes as "habits" that can be effectively intervened through autonomous identification, regulation, and systematic behaviors. Its connotation aligns with UBT, including both education and training components (Devine et al., 2012; Forscher & Devine, 2017). Research demonstrates that PHBI is an effective method for reducing implicit bias and has long-term effects. For example, Devine et al. (2012) conducted a three-month longitudinal study and found that the experimental group receiving PHBI intervention showed significantly reduced implicit racial bias and significantly increased concern about discrimination, with effects persisting until week 8, while the control group showed no similar changes (Devine et al., 2012). PHBI also functions in real-world social contexts (Cox & Devine, 2019). For instance, multiple studies show that PHBI interventions in STEM (Science, Technology, Engineering, and Mathematics) fields reduce implicit gender bias and stereotypes, significantly increase female hiring rates, and promote gender diversity and social equity (Carnes et al., 2015, 2023; Devine et al., 2017). In the latest research, PHBI was applied to implicit bias training for Swedish social workers. Results showed that although the reduction in implicit bias in the intervention group did not achieve expected effects, participants' awareness of implicit bias significantly increased (Bursell, 2024). This finding suggests that PHBI' s effects may be influenced by specific social and cultural contexts, emphasizing the need to consider contextual factors when applying interventions.

4.5.3 Priming Multiculturalism Priming multiculturalism reduces implicit bias toward specific groups by encouraging individuals to recognize and appreciate racial and cultural diversity. Participants are guided to read and reflect on multiculturalism concepts, followed by IAT measurement of implicit racial

bias. Results show that priming multiculturalism effectively reduces implicit racial bias (FitzGerald et al., 2019; Lai et al., 2016; Lai et al., 2014). Sadler et al. examined the impact of racial diversity in 351 U.S. metropolitan areas on implicit bias (such as implicit associations between Black people and weapons). The study found that increasing racial diversity and integration effectively reduces implicit bias toward Black people (Sadler & Devos, 2020). Individuals who participated in diversity training (including contact, communication, and promoting social identity) also showed reduced implicit bias (Behm-Morawitz & Villamil, 2019). This indicates that promoting racial and group diversity and integration through diversity training is more effective in reducing implicit bias than simply increasing minority group proportions.

In summary, as a comprehensive emerging intervention method, UBT can help people enhance bias awareness, guide them in identifying and managing bias, and thus more consciously avoid generating bias and its negative consequences in daily life (Bezyak et al., 2024; Cooper et al., 2022; Howell et al., 2024; Vitriol & Banaji, 2024). Most studies reveal UBT' s effectiveness (Liu et al., 2022; Stone et al., 2020; Liang et al., 2019; Jackson et al., 2014), though some research finds UBT ineffective or even counterproductive (Williamson & Foley, 2018). For example, Lai' s study examined whether short-term diversity training for implicit bias could reduce police officers' implicit bias and found that although the training increased police awareness of bias, it did not effectively reduce their implicit bias (Lai & Lisnek, 2023). UBT' s durability also requires continuous external reinforcement and consolidation to maintain (Lindvall-Östling, 2024).

4.6 Implementation Intentions

Implementation intentions are strategies based on "if-then" logical structures that link intended goal behaviors with specific situational cues to achieve a conditional reflex form of goal implementation (Sheeran et al., 2024). Lai et al. (2014) explored the role of implementation intentions in reducing implicit racial bias. In the experiment, White participants were trained to form specific implementation intentions, such as "If I encounter a Black face, then I will immediately have a positive reaction." Results showed this method significantly reduced implicit racial bias in the short term, though its long-term effects have not been verified (Calanchini et al., 2021; Lai et al., 2016). Additionally, implementation intentions show effectiveness in reducing implicit stereotypes (FitzGerald et al., 2019; Rees et al., 2019; B. D. Stewart & Payne, 2008). For example, Mendoza (2010) tested the impact of two types of implementation intentions—distraction inhibition and response facilitation—on stereotype task accuracy through two studies. Results showed both implementation intentions improved task performance accuracy, i.e., suppressed implicit stereotype behavioral expression, and this effect was not achieved by prolonging task reaction time, indicating that implementation intentions function through reflexive control rather than increased reaction time (Mendoza et al., 2010). In summary, implementation intentions form an automatic conditional reflex by linking target behaviors with specific

situational cues, helping individuals automatically generate positive responses through reflexive control, thereby reducing implicit bias, implicit stereotypes, and corresponding behaviors in the short term.

4.7 Targeted Memory Reactivation (TMR)

Sleep plays an important role in strengthening and consolidating memories acquired during wakefulness. This spontaneous memory reactivation process during sleep can be manipulated through Targeted Memory Reactivation (TMR) technology. TMR enhances memory consolidation effects by presenting sensory cues (such as specific sounds or smells) related to previous learning experiences during sleep (Carbone & Diekelmann, 2024). Hu et al. first applied TMR to implicit bias intervention and found that participants who received "counter-stereotype training" —consciously constructing associations contradicting stereotypes to form counter-stereotypical associations—and subsequent TMR showed significantly reduced implicit bias with effects lasting at least one week (Hu et al., 2015), emphasizing the importance of memory consolidation in long-term intervention effects. However, Humiston et al.' s study failed to replicate Hu et al.' s results, finding no significant effect of TMR on implicit bias whether measured immediately after a nap or one week later (Humiston & Wamsley, 2019). Based on this, researchers emphasized the need for more sensitive analytical methods to verify TMR' s long-term effectiveness in such interventions (Aczel et al., 2015). In response to inconsistent findings, Hu' s team reanalyzed data and found that intervening during specific sleep stages, particularly implementing TMR during the slow oscillation up-state, can more effectively reduce implicit social bias, with effects persisting in tests one week later (Xia et al., 2022). Additionally, Chen et al.' s latest research expanded TMR' s application by exploring the combination of self-esteem cognitive training with TMR to enhance implicit self-esteem. Results showed that using specific auditory cues to activate memories during sleep can effectively promote long-term improvement in implicit self-esteem (Chen et al., 2021). In summary, TMR is an effective implicit social cognition intervention strategy with long-term durability, though more research is needed to further demonstrate the robustness of intervention effects.

4.8 Emotion Induction

Implicit bias is largely influenced by emotions. Research shows that emotional states can modulate the activation and expression of individuals' implicit stereotypes. Specifically, negative emotions can reduce the expression of implicit stereotypes and attitudes compared to positive emotions (Huntsinger et al., 2009). However, in specific contexts, such as when individuals hold counter-stereotypical thinking or egalitarian attitudes, positive emotions can also effectively reduce the activation of negative implicit social cognition (Huntsinger et al., 2010). Other research indicates that emotional ability moderates the formation and change of implicit attitudes by influencing evaluative conditioning

(Hasford et al., 2018). However, inducing emotions such as guilt or moral sense is difficult for effectively reducing individuals' implicit racial bias (Lai et al., 2016; Lai et al., 2014).

Emotion induction interventions have corresponding neural bases. For example, White individuals show more significant amygdala activation when viewing Black faces (Lieberman et al., 2005). Based on this, researchers have explored pharmacological neural interventions for implicit bias. For instance, drugs can reduce the amygdala' s response intensity to emotional stimuli (Hurlemann et al., 2010). Accordingly, Terbeck found that pharmacological intervention using propranolol (a non-selective β-receptor blocker) significantly reduced implicit racial bias (compared to a placebo group) (Terbeck et al., 2012, 2015).

4.9 Mindfulness Meditation

Mindfulness Meditation is a meditative practice that enhances individuals'awareness and acceptance of present-moment experiences by cultivating a "mindful" state. In this process, individuals focus attention on current experiences and observe their feelings, thoughts, and emotions with a non-judgmental, open, and curious attitude. This meditation practice has been widely applied in various psychological fields, particularly showing significant effects in promoting mental health, reducing stress, and improving emotion regulation (Ngan & Cheng, 2022). Research finds that mindfulness meditation plays a role in changing implicit social cognition, such as reducing implicit stereotypes and group bias and enhancing individuals' implicit self-esteem (Djikic et al., 2008; Kang et al., 2013; Narayan, 2019). Mindfulness meditation can also reduce automatic biased thinking, thereby promoting prosocial behavior. Lueke et al. found that brief mindfulness meditation audio interventions can reduce implicit racial and gender bias (Hunsinger et al., 2019; Lueke & Gibson, 2015).

Mindfulness meditation has broad applications in education, workplace, and clinical contexts. Recent research shows that teachers' mindfulness levels moderate their implicit bias toward Black students when dealing with educational situations involving Black students. Teachers with higher mindfulness levels show less racial bias toward Black students, while teachers with lower mindfulness levels may exhibit more bias when handling issues involving Black students (Ash et al., 2023; Hirshberg et al., 2022). This result suggests that improving teachers' mindfulness levels can help reduce biased decision-making and thus reduce educational inequity. Further research found that mindfulness interventions can improve teachers' mindfulness levels and significantly reduce their implicit bias toward Black students (Romano, 2024). However, in workplace contexts, mindfulness meditation interventions to reduce implicit age bias were not significant (Williams & Polito, 2022). This indicates that single mindfulness meditation training' s intervention effects on reducing implicit bias remain unstable.

5 Summary and Outlook

Implicit social cognition has important impacts on individual psychology and behavior and even on macro-level society (Greenwald & Lai, 2020; Yang et al., 2015). Implicit social cognition is stable yet also influenced by situational factors, showing certain variability (Kurdi & Charlesworth, 2023; Rauf et al., 2022). Based on this, researchers have recently dedicated efforts to exploring methods for intervening in implicit social cognition to reduce negative implicit social cognition (such as implicit bias, implicit stereotypes, low implicit self-esteem) and their resulting negative individual and social impacts. This article integrated and analyzed existing research on implicit social cognition interventions, summarizing nine intervention methods. According to their definitions and principles, these methods can be divided into four categories: (1) **Cognitive Reconstruction**: changing individuals' cognitive or associative patterns to intervene in implicit social cognition, such as evaluative conditioning, approach-avoidance training, exposure to counter-stereotypical exemplars, unconscious bias training, and implementation intentions; (2) **Social Interaction**: reducing bias through interaction with individuals or groups, such as intergroup contact; (3) **Memory Enhancement**: intervening in implicit social cognition through memory reactivation or consolidation, such as targeted memory reactivation; (4) **Emotion Regulation**: intervening in implicit social cognition by regulating individuals' emotional or mood states, such as emotion induction and mindfulness meditation.

All these methods demonstrate certain effectiveness, but their utility varies significantly across different studies and intervention contents. This article conducted meta-analyses of effect sizes (based on Cohen' s d) for each method (Table 1). Results show that these nine intervention methods have small to medium effect sizes (0.32~0.58), with approach-avoidance training, unconscious bias training, and emotion induction interventions being relatively effective, showing medium effect sizes. Some methods, such as implementation intentions and targeted memory reactivation, have fewer studies, so the reliability of meta-analysis results requires further verification in future research. It should be noted that this article focuses on analyzing empirically validated effective intervention methods and does not introduce methods that are ineffective when used alone, such as perspective-taking and empathy training (Calanchini et al., 2021; Skorinko et al., 2023). Research shows that combining multiple methods can effectively improve intervention effectiveness. For example, Gabrielli et al. combined intergroup contact strategies with empathy training and perspective-taking and found that children receiving multi-strategy interventions showed significantly reduced implicit bias toward immigrants (Gabrielli et al., 2022). Additionally, combining mindfulness meditation with interpersonal relationship promotion methods can effectively reduce preservice teachers' implicit racial bias, with effects still present at six-month follow-up (Hirshberg et al., 2022).

Although the above intervention methods have achieved certain short-term suc-

cess, many methods' durability is typically limited to the short term and have not shown significant long-term utility, such as approach-avoidance training and intergroup contact. Some methods even show counterproductive effects in certain contexts, such as unconscious bias training and negative intergroup contact (Williamson & Foley, 2018). Thus, the durability of intervention effects has become an important issue for existing methods, mainly attributable to three reasons: First, most intervention methods rely on laboratory experiments, ignoring contextual factors and limiting generalizability to real-world settings; second, existing interventions are typically too singular with insufficient sessions, failing to create continuous reinforcement effects; third, existing methods primarily target changing cognition itself while ignoring sociocultural factors in the formation process of implicit social cognition.

Research shows that interventions with strong durability typically have highly contextualized characteristics and require long-term, repeated, or multi-method combinations within specific target populations to achieve optimal effects (Chapman et al., 2018; Devine et al., 2017; Stone et al., 2020). Therefore, future intervention research could focus on three directions: First, develop more targeted and contextualized intervention strategies to ensure ecological validity; second, adopt multi-method combinations with repeated reinforcement interventions to improve long-term effects; third, explore the influence of cultural and social backgrounds on implicit social cognition to clarify intervention applicability boundaries and ensure intervention suitability.

With the development of artificial intelligence technology, implicit social cognition intervention methods can expand from traditional laboratory environments to more ecological application scenarios. For example, intervention methods based on Virtual Reality (VR) and natural language processing can advance future implicit social cognition interventions. VR can simulate realistic, immersive social scenarios where individuals can experience different social roles and identities and perceive diverse social interactions and cultural differences (Marini & Casile, 2022; Patané et al., 2020). VR interventions can also be personalized, setting scenarios based on individuals' implicit bias characteristics to enhance immersive intervention effects (Shen et al., 2024). This highly immersive, personalized experience can effectively intervene in implicit social cognition (Tassinari et al., 2022, 2024). VR technology is also effective in identifying implicit bias, as Carson et al. identified implicit group bias in public reporting through VR technology, which could inform targeted corrections to reduce implicit bias (Carson & Politte, 2021).

Word embeddings, as a key technology in natural language processing, have important application value in revealing and intervening in implicit social cognition. Word vectors trained on large-scale corpora can reflect societal implicit attitudes and stereotypes regarding race, gender, and other dimensions (Bao et al., 2023). For example, word embedding-based analyses have revealed the existence of implicit bias in natural language, showing that White/Black groups are semantically closer to attribute words like "superior"or "inferior"(Lee et al., 2024). Future

research could use word embedding technology to identify biased tendencies and expressions in online platforms and guide users through appropriate feedback mechanisms to intervene in implicit bias. For instance, Kroon identified implicit bias in news media reporting through word embeddings, emphasizing media' s responsibility to reduce unconscious bias to avoid unfair stereotypes of specific ethnic groups (Kroon & van der Meer, 2023). Future research could further develop and promote these natural language processing-based tools to automatically identify implicit bias in large-scale data environments and combine them with various AI technologies such as large language models, affective computing, and machine learning to form intelligent, personalized intervention systems. For example, using natural language processing technology to analyze individuals' online texts to identify embedded biases and emotional tendencies, combined with machine learning algorithms for real-time monitoring and dynamic analysis of individuals' texts to identify cognitive changes in different contexts and detect potential social cognition problems in a timely manner (Zhang et al., 2025; Ji et al., 2020). Large language models can generate personalized intervention content based on identified information, such as targeted suggestions, persuasive messages, or specific action plans (Huang et al., 2024). Through these technologies, individuals' implicit social cognition in different contexts can be identified, appropriate implicit training programs can be automatically generated, intervention strategies can be continuously optimized, and real-time, dynamic intervention suggestions can be provided to ensure the durability of intervention effects, thereby achieving real-time personalized intervention for individual and group behaviors through AI technology.

Furthermore, macro-level education and culture play important roles in reducing negative implicit social cognition. Education can cultivate individuals' multicultural awareness through targeted curriculum design and teaching activities, thereby reducing implicit bias and stereotypes. At the cultural level, promoting cross-cultural communication and cultural diversity education can enhance individuals' understanding and tolerance of different cultures, further weakening negative implicit social cognition. This system integrating technology, education, and cultural intervention can not only provide real-time, convenient interventions for people but also be widely applied in practical scenarios such as education, healthcare, and corporate training, thereby achieving effective intervention in implicit social cognition at both individual and societal levels.

(*References marked with asterisks were included in the meta-analysis)

References

Bao, H., Wang, Z., Cheng, X., Su, Z., Yang, Y., Zhang, G., ···& Cai, H. (2023). Psychological research based on word embedding technology: Methods and applications. *Advances in Psychological Science*, 31(6), 887.

*Shi, T. (2007). An experimental study on changing implicit self-esteem through evaluative conditioning [Master' s thesis]. Yunnan Normal University.

*Chen, J. (2016). A study on college students' implicit attitudes toward people with disabilities [Master' s thesis]. Zhengzhou University.

*Cao, R. (2015). The impact of imagined intergroup contact on explicit and implicit bias toward gay men [Master' s thesis]. Sichuan Normal University.

Gao, K., Zhang, Y., Li, S., Yuan, J., Li, H., & Zhang, D. (2023). The causal role of ventromedial prefrontal cortex in implicit cognitive reappraisal. *Acta Psychologica Sinica*, 55(2), 210.

*Xie, S. (2019). The impact of imagined intergroup contact on implicit bias toward obese women [Master' s thesis]. Shaanxi Normal University.

*Liu, Y. (2010). Explicit and implicit attitudes of heterosexual college students toward homosexuality and their intervention [Master' s thesis]. Zhengzhou University.

Liu, L., Zhao, Y., Huang, Y., Wang, Y., & Zhu, Y. (2012). Enhancing implicit self-esteem to reduce aggression induced by social exclusion. Journal of Southwest China Normal University (Natural Science Edition)*, (8), 135-138.

*Li, Y. (2012). The impact of imagined intergroup contact and contact experience on implicit bias toward migrant workers [Master' s thesis]. Soochow University.

Yang, Z., Liu, Y., Yu, Z., & Cai, H. (2015). Applications of the Implicit Association Test: Current research status at home and abroad. *Advances in Psychological Science*, 23(11), 1879-1890.

Zhu, R., & Chen, C. (2013). The impact of evaluative conditioning on implicit attitudes, explicit attitudes, and behaviors toward healthy eating. Chinese Ergonomics*, 19(2), 66-70.

Zhang, Y., Huang, F., Mo, L., Liu, X., & Zhu, T. (in press). (2025). Suicidal ideation data augmentation and identification technology based on large language models. *Acta Psychologica Sinica*.

Aczel, B., Palfi, B., Szaszi, B., Szollosi, A., & Dienes, Z. (2015). Commentary: Unlearning implicit social biases during sleep. *Frontiers in Psychology*, 6, Article 1428. https://doi.org/10.3389/fpsyg.2015.01428

Ash, T. L., Helminen, E. C., Yamashita, S., & Felver, J. C. (2023). Teachers' anti-black biases in disciplinary decisions: The role of mindfulness. *Journal of School Psychology*, 96, 75-89.

Aberson, C. L., Shoemaker, C., & Tomolillo, C. (2004). Implicit bias and contact: The role of interethnic friendships. The Journal of Social Psychology*, 144(3), 335-347.

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science*, 15(7), 498-502.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., ···& Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, 38(12), 1629-1643.

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives*, 9(1), 50-54.

Behm-Morawitz, E., & Villamil, A. M. (2019). The roles of ingroup identification and implicit bias in assessing the effectiveness of an online diversity education program. Journal of Applied Communication Research*, 47(5), 505-526.

Bezyak, J., Versen, E., Chan, F., Lee, D., Wu, J. R., Iwanaga, K., ···& Ho, H. (2024). Needs of human resource professionals in implicit bias and disability inclusion training: A focus group study. *Journal of Vocational Rehabilitation*, 60(3), 311-319.

Bond, J., Dixon, J., Tredoux, C., & Andreouli, E. (2023). The contact hypothesis and the virtual revolution: Does face-to-face interaction remain central to improving intergroup relations? *PLOS ONE*, 18(12), Article e0292831. https://doi.org/10.1371/journal.pone.0292831

Breves, P. (2020). Reducing outgroup bias through intergroup contact with non-playable video game characters in VR. Presence*, 27(3), 257-273.

Bursell, M. (2024). The scope and limits of implicit bias training: An experimental study with Swedish social workers. Social Science Information*, 63(1), 92-113.

Crescentini, C., Urgesi, C., Campanella, F., Eleopra, R., & Fabbro, F. (2014). Effects of an 8-week meditation program on the implicit and explicit attitudes toward religious/spiritual self-representations. Consciousness and Cognition*, 30, 266-280.

Castillo, L. G., Brossart, D. F., Reyes, C. J., Conoley, C. W., & Phoummarath, M. J. (2007). The influence of multicultural training on perceived multicultural counseling competencies and implicit racial prejudice. *Journal of Multicultural Counseling and Development*, 35(4), 243-255.

Columb, C., & Plant, E. A. (2011). Revisiting the Obama effect: Exposure to Obama reduces implicit prejudice. *Journal of Experimental Social Psychology*, 47(2), 499-501.

Cai, H., Luo, Y. L. L., Chen, J., Li, X., & Xie, Y. (2021). Is implicit social cognition developmentally stable? A longitudinal study. *Developmental Psychology*, 57(12), 2220-2235.

Cai, H., & Wu, L. (2021). The self-esteem Implicit Association Test is valid: Evidence from brain activity. *PsyCh Journal*, 10(3), 465-476.

Calanchini, J., Lai, C. K., & Klauer, K. C. (2021). Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *Journal of Personality and Social Psychology*, 121(4), 796-815.

Choi, Y. J., & Lee, J. H. (2015). Alcohol-related attitudes of heavy drinkers: Effects of arousal and valence in evaluative conditioning. Social Behavior and Personality: An International Journal*, 43(2), 205-217.

Carbone, J., & Diekelmann, S. (2024). An update on recent advances in targeted memory reactivation during sleep. *Npj Science of Learning*, 9(1), Article 31. https://doi.org/10.1038/s41539-024-00244-8

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., ···& Sheridan, J. (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), 221-230.

Carnes, M., Sheridan, J., Fine, E., Lee, Y. G., & Filut, A. (2023). Effect of a workshop to break the bias habit for internal medicine faculty: A multisite cluster randomized controlled study. *Academic Medicine*, 98(10), 1211-1219.

Carson, J. V., & Politte, H. (2021). Implicit bias within public reporting: A virtual reality experiment examining "suspicious"activity. *Crime & Delinquency*, 67(12), 2135-2162.

Chapman, M. V., Hall, W. J., Lee, K., Colby, R., Coyne-Beasley, T., Day, S., ··· & Payne, K. (2018). Making a difference in medical trainees' attitudes toward Latino patients: A pilot study of an intervention to modify implicit and explicit attitudes. *Social Science & Medicine*, 199, 202-209.

Charlesworth, T. E., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, 30(2), 174-192.

Charlesworth, T. E., & Banaji, M. R. (2022a). Patterns of implicit and explicit attitudes: IV. Change and stability from 2007 to 2020. *Psychological Science*, 33(9), 1347-1363.

Charlesworth, T. E., & Banaji, M. R. (2021). Patterns of implicit and explicit attitudes II. Long-term change and stability, regardless of group membership. *American Psychologist*, 76(6), 851-868.

Charlesworth, T. E., & Banaji, M. R. (2022b). Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Social Psychological and Personality Science*, 13(1), 14-26.

Charlesworth, T. E., Kurdi, B., & Banaji, M. R. (2020). Children' s implicit attitude acquisition: Evaluative statements succeed, repeated pairings fail. *Developmental Science*, 23(3), Article e12911. https://doi.org/10.1111/desc.12911

Chen, L., Wang, Y., Zhang, Y., & Wang, Y. (2023). A new application of TMR: A study on implicit self-esteem. Current Psychology*, 42(7), 5159-5168.

Cooper, L. A., Saha, S., & Van Ryn, M. (2022, August). Mandated implicit bias training for health professionals—a step toward equity in health care. *JAMA Health Forum*, 3(8), e223250. https://doi.org/10.1001/jamahealthforum.2022.3250

Chang, Y. (2018). Exploring the patterns of dual attitude changes in the context of athlete endorsement: The interplay of fit, evaluative conditioning, and introspection focus. Journal of Sport Management*, 32(5), 413-425.

Cox, W. T., & Devine, P. G. (2019). The prejudice habit-breaking intervention: An empowerment-based confrontation approach. In *Confronting Prejudice and Discrimination* (pp. 249-274). Academic Press.

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math-gender stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, 117, 73-91.

Cvencek, D., Nasir, N. I. S., O' Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The development of math-race stereotypes: "They say Chinese people are the best at math" . *Journal of Research on Adolescence*, 25(4), 630-637.

Critcher, C. R., & Risen, J. L. (2014). If he can do it, so can they: Exposure to counterstereotypically successful exemplars prompts automatic inferences. Journal of Personality and Social Psychology*, 106(3), 359-376.

Dasgupta, N., & Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. Social Cognition*, 26(1), 112-123.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. Journal of Personality and Social Psychology*, 81(5), 800-814.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 1267-1278.

Devine, P. G., & Ash, T. L. (2022). Diversity training goals, limitations, and promise: A review of the multidisciplinary literature. *Annual Review of Psychology*, 73(1), 403-429.

Devine, P. G., Forscher, P. S., Cox, W. T., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. Journal of Experimental Social Psychology*, 73, 211-215.

Dijksterhuis, A. P. (2004). I like myself but I don' t know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. Journal of Personality and Social Psychology*, 86(2), 345-355.

Djikic, M., Langer, E. J., & Stapleton, S. F. (2008). Reducing stereotyping

through mindfulness: Effects on automatic stereotype-activated behaviors. Journal of Adult Development*, 15, 106-111.

Di Bernardo, G. A., Vezzali, L., Birtel, M. D., Stathi, S., Ferrari, B., Giovannini, D., & Pettigrew, T. F. (2022). The role of optimal conditions and intergroup contact in promoting positive intergroup relations in and out of the workplace: A study with ethnic majority and minority workers. *Group Processes & Intergroup Relations*, 25(6), 1516-1533.

Doan, L., Ray, R., Powelson, C., Fuentes, G., Shankman, R., Genter, S., & Bailey, J. (2021, July). Evaluation of a virtual reality simulation tool for studying bias in police-civilian interactions. In *International Conference on Human-Computer Interaction* (pp. 388-399). Springer International Publishing.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, 12(7), 248-253.

Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological Science*, 24(6), 860-868.

Duque, M., De Coninck, D., Montero-Zamora, P., Sahbaz, S., Bautista, T., Ertanir, B., ···& Schwartz, S. J. (2024). The spillover effect of implicit immigrant stereotypes and incidental pandemic threat. *International Journal of Intercultural Relations*, 99, Article 101939. https://doi.org/10.1016/j.ijintrel.2024.101939

Eisenstadt, B. E., Spagna, A., & Stroessner, S. J. (2024). In the Mind' s Eye: Exploring the Relationship Between Visual Mental Imagery and Stereotyping. *Personality and Social Psychology Review*. https://doi.org/10.1177/10888683241273350

Ebert, I. D., Steffens, M. C., Von Stülpnagel, R., & Jelenec, P. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. Journal of Experimental Social Psychology*, 45(5), 1098-1104.

Falk, C. F., & Heine, S. J. (2015). What is implicit self-esteem, and does it vary across cultures? *Personality and Social Psychology Review*, 19(2), 177-196.

Farmer, H., Hewstone, M., Spiegler, O., Morse, H., Saifullah, A., Pan, X., ··· & Terbeck, T. (2020). Positive intergroup contact modulates fusiform gyrus activity to black and white faces. *Scientific Reports*, 10(1), Article 2700. https://doi.org/10.1038/s41598-020-59633-9

FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology*, 7, 1-12.

Forscher, P. S., & Devine, P. G. (2014). Breaking the prejudice habit: Automaticity and control in the context of a long-term goal. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-Process Theories of the Social Mind* (pp. 468-482). The Guilford Press.

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing Muslim/Arab stereotypes through evaluative conditioning. The Journal of Social Psychology*, 153(1), 6-9.

Freud, S. (1960). *Jokes and Their Relation to the Unconscious.* New York: WW Norton & Company.

Gabrielli, S., Catalano, M. G., Maricchiolo, F., Paolini, D., & Perucchini, P. (2022). Reducing implicit prejudice towards migrants in fifth grade pupils: Efficacy of a multi-faceted school-based program. Social Psychology of Education*, 25(2), 425-445.

Gawronski, B. (2024). Implicit bias: What is it and how does it matter for diversity, equity, and inclusion? In V. M. Esses, J. F. Dovidio, J. Jetten, D. Sekaquaptewa, & K. West (Eds.), *The Sage Handbook of Psychological Perspectives on Diversity, Equity, and Inclusion* (pp. 71-168). Sage.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. Journal of Experimental Social Psychology*, 44(2), 370-377.

Gawronski, B., Ledgerwood, A., & Eastwick, P. W. (2022). Implicit Bias ≠ Bias on Implicit Measures. *Psychological Inquiry*, 33(3), 139-150.

Gonzalez, A. M., Steele, J. R., Chan, E. F., Lim, S. A., & Baron, A. S. (2021). Developmental differences in the malleability of implicit racial bias following exposure to counterstereotypical exemplars. Developmental Psychology*, 57(1), 102-112.

Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing children's implicit racial bias through exposure to positive out-group exemplars. Child Development*, 88(1), 123-137.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109(1), 3-25.

Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friese, M., ···& Wiers, R. W. (2022). Best research practices for using the Implicit Association Test. *Behavior Research Methods*, 54(3), 1115-1135.

Greenwald, A. G., Dasgupta, N., Dovidio, J. F., Kang, J., Moss-Racusin, C. A., & Teachman, B. A. (2022). Implicit-bias remedies: Treating discriminatory bias as a public-health problem. *Psychological Science in the Public Interest*, 23(1), 7-38.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal*

of Personality and Social Psychology, 74(6), 1464-1480.

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, 71(1), 419-445.

Grumm, M., Nestler, S., & Von Collani, G. (2009). Changing explicit and implicit attitudes: The case of self-esteem. Journal of Experimental Social Psychology*, 45(2), 327-335.

Geng, L., Liu, L., Xu, J., Zhou, K., & Fang, Y. (2013). Can evaluative conditioning change implicit attitudes towards recycling? Social Behavior and Personality: An International Journal*, 41(6), 947-955.

Hollands, G. J., & Marteau, T. M. (2016). Pairing images of unhealthy and healthy foods with images of negative and positive health consequences: Impact on attitudes and food choice. Health Psychology*, 35(8), 847-857.

Houben, K., Havermans, R. C., & Wiers, R. W. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes toward alcohol and its effect on drinking behavior. Psychopharmacology*, 211, 79-86.

Hagiwara, N., Duffy, C., Cyrus, J., Harika, N., Watson, G. S., & Green, T. L. (2024). The nature and validity of implicit bias training for health care providers and trainees: A systematic review. *Science Advances*, 10(33), Article eado5957. https://doi.org/10.1126/sciadv.ado5957

Hasford, J., Kidwell, B., & Hardesty, D. M. (2018). Emotional ability and associative learning: How experiencing and reasoning about emotions impacts evaluative conditioning. *Journal of Consumer Research*, 45(4), 743-761.

Hirshberg, M. J., Flook, L., Moss, E. E., Enright, R. D., & Davidson, R. J. (2022). Integrating mindfulness and connection practices into preservice teacher education results in durable automatic race bias reductions. Journal of School Psychology*, 91, 50-65.

Huang, F., Sun, X., Mei, A., Wang, Y., Ding, H., & Zhu, T. (2024). LLM Plus Machine Learning Outperform Expert Rating to Predict Life Satisfaction From Self-Statement Text. *IEEE Transactions on Computational Social Systems*. https://doi.org/10.1109/TCSS.2024.3475413

Howell, J. L., Lofaro, N., & Ratliff, K. A. (2024). Responding to feedback about implicit bias. *Social and Personality Psychology Compass*, 18(1), Article e12926. https://doi.org/10.1111/spc3.12926

Hu, X., Antony, J. W., Creery, J. D., Vargas, I. M., Bodenhausen, G. V., & Paller, K. A. (2015). Unlearning implicit social biases during sleep. Science*, 348(6238), 1013-1015.

Humiston, G. B., & Wamsley, E. J. (2019). Unlearning implicit social biases during sleep: A failure to replicate. PLOS ONE*, 14(1), Article e0211416. https://doi.org/10.1371/journal.pone.0211416

Hunsinger, M., Christopher, M., & Schmidt, A. M. (2019). Mindfulness training, implicit bias, and force response decision-making. Mindfulness*, 10(12), 2555-2565.

Huntsinger, J. R., Sinclair, S., & Clore, G. L. (2009). Affective regulation of implicitly measured stereotypes and attitudes: Automatic and controlled processes. Journal of Experimental Social Psychology*, 45(3), 560-566.

Huntsinger, J. R., & Smith, C. T. (2009). First thought, best thought: Positive mood maintains and negative mood degrades implicit-explicit attitude correspondence. Personality and Social Psychology Bulletin*, 35(2), 187-197.

Huntsinger, J. R., Sinclair, S., Dunn, E., & Clore, G. L. (2010). Affective regulation of stereotype activation: It' s the (accessible) thought that counts. Personality and Social Psychology Bulletin*, 36(4), 564-577.

Huntsinger, J. R., & Sinclair, S. (2010). When it feels right, go with it: Affective regulation of affiliative social tuning. Social Cognition*, 28(3), 290-305.

Huntsinger, J. R. (2011). Mood and trust in intuition interactively orchestrate correspondence between implicit and explicit attitudes. Personality and Social Psychology Bulletin*, 37(9), 1245-1258.

Huntsinger, J. R. (2013). Anger enhances correspondence between implicit and explicit attitudes. Emotion*, 13(2), 350-356.

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., ···& Kendrick, K. M. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *Journal of Neuroscience*, 30(14), 4999-5007.

Jackson, S. M., Hillard, A. L., & Schneider, T. R. (2014). Using implicit bias training to improve attitudes toward women in STEM. Social Psychology of Education*, 17(3), 419-438.

Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. Social Psychology*, 41(3), 137-146.

Jeon, B., Lee, M., & Lee, J. H. (2019). Evaluative conditioning-induced implicit attitude change in violent juvenile offenders: A randomized controlled trial. Psychology of Violence*, 9(1), 28-38.

Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, 8(1), 214-226.

Jin, Z., Shiomura, K., Bermeitinger, C., Zhu, X., De Paula Couto, M. C. P., Wu, L., & Baess, P. (2022). Twenty Years After Implicit Association Test: The Role of Implicit Social Cognition in Human Behavior. *Frontiers in Psychology*, 13, Article 919658. https://doi.org/10.3389/fpsyg.2022.919658

Kang, Y., Gray, J. R., & Dovidio, J. F. (2013). The nondiscriminating heart: Loving-kindness meditation training decreases implicit intergroup bias. Journal of Experimental Psychology: General*, 143(3), 1306-1313.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. Journal of Personality and Social Psychology*, 92(6), 957-971.

Kroon, A. C., & van der Meer, T. G. (2023). Who' s to fear? Implicit sexual threat pre and post the "refugee crisis" . Journalism Practice*, 17(2), 319-335.

Kurdi, B., & Charlesworth, T. E. (2023). A 3D framework of implicit attitude change. *Trends in Cognitive Sciences*, 27(8), 745-758.

Kurdi, B., Mann, T. C., Charlesworth, T. E., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences*, 116(13), 5862-5871.

Kang, Y., & Falk, E. B. (2020). Neural mechanisms of attitude change toward stigmatized individuals: Temporoparietal junction activity predicts bias reduction. Mindfulness*, 11, 1378-1389.

Liao, T., Zhuoga, C., & Chen, X. (2023). Contact with grandparents and young people' s explicit and implicit attitudes toward older adults. BMC Psychology*, 11(1), Article 289. https://doi.org/10.1186/s40359-023-01345-8

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, 7(5), 315-330.

Lai, C. K., & Lisnek, J. A. (2023). The impact of implicit-bias-oriented diversity training on police officers' beliefs, motivations, and actions. *Psychological Science*, 34(4), 424-437.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., ···& Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. Journal of Experimental Psychology: General*, 143(4), 1765-1785.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ···& Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. Journal of Experimental Psychology: General*, 145(8), 1001-1016.

Lee, M. H., Montgomery, J. M., & Lai, C. K. (2024). America's racial framework of superiority and Americanness embedded in natural language. *PNAS Nexus*, 3(1), pgad485.

Levontin, L., Halperin, E., & Dweck, C. S. (2013). Implicit theories block negative attributions about a longstanding adversary: The case of Israelis and Arabs. Journal of Experimental Social Psychology*, 49(4), 670-677.

Liang, J., Wolsiefer, K., Zestcott, C. A., Chase, D., & Stone, J. (2019). Implicit bias toward cervical cancer: Provider and training differences. *Gynecologic Oncology*, 153(1), 80-86.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, 8(6), 720-722.

Lindvall-Östling, M. (2024). "If you are aware of something, you can do something about it" : Investigating the lasting effects of an implicit bias training activity. *Teaching and Teacher Education*, 140, Article 104484. https://doi.org/10.1016/j.tate.2024.104484

Liu, F. F., Coifman, J., McRee, E., Stone, J., Law, A., Gaias, L., ···& Lyon, A. R. (2022). A brief online implicit bias intervention for school mental health clinicians. International Journal of Environmental Research and Public Health*, 19(2), Article 679. https://doi.org/10.3390/ijerph19020679

Lueke, A., & Gibson, B. (2015). Mindfulness meditation reduces implicit age and race bias: The role of reduced automaticity of responding. Social Psychological and Personality Science*, 6(3), 284-291.

Luo, C., He, M., & Xing, C. (2024). Public acceptance of autonomous vehicles in China. *International Journal of Human-Computer Interaction*, 40(2), 315-327.

Marini, M., & Casile, A. (2022). I can see my virtual body in a mirror: The role of visual perspective in changing implicit racial attitudes using virtual reality. *Frontiers in Psychology*, 13, Article 989582. https://doi.org/10.3389/fpsyg.2022.989582

Marini, M., Rubichi, S., & Sartori, G. (2011). Implicit racial stereotypes may temporarily shift after reading a story. Cognitive Research: Principles and Implications*, 6. https://doi.org/10.13140/2.1.1888.7688

Marini, M., Sabin, J., O' shea, B., & Vianello, M. (2024). Implicit social cognition: Malleability and change. *Frontiers in Psychology*, 15, Article 1475986. https://doi.org/10.3389/fpsyg.2024.1475986

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. Personality and Social Psychology Bulletin*, 36(4), 512-523.

Möller, C., Passam, S., Riley, S., & Robson, M. (2024). All inside our heads? A critical discursive review of unconscious bias training in the sciences. *Gender, Work and Organization*, 31(3), 797-815.

Moran, T., Nudler, Y., & Bar-Anan, Y. (2023). Evaluative conditioning: Past, present, and future. *Annual Review of Psychology*, 74(1), 245-270.

Narayan, M. C. (2019). CE: Addressing implicit bias in nursing: A review. *AJN The American Journal of Nursing*, 119(7), 36-42.

Ngan, S. T. J., & Cheng, P. W. C. (2022). Classification of mindfulness meditation and its impact on neural measures in the clinical population. *Frontiers in Psychology*, 13, Article 891004. https://doi.org/10.3389/fpsyg.2022.891004

*Ngnoumen, C. T. (2019). The use of socio-cognitive mindfulness in mitigating implicit bias and stereotype-activated behaviors. Harvard University.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231-259.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. Personality and Social Psychology Bulletin*, 32(4), 421-433.

Onyeador, I. N., Wittlin, N. M., Burke, S. E., Dovidio, J. F., Perry, S. P., Hardeman, R. R., ···& van Ryn, M. (2020). The value of interracial contact for reducing anti-black bias among non-black physicians: A cognitive habits and growth evaluation (CHANGE) study report. Psychological Science*, 31(1), 18-30.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2), 129-158.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, 72(1), 533-560.

Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes toward Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology*, 29(1), 35-45.

Patané, I., Lelgouarch, A., Banakou, D., Verdelet, G., Desoche, C., Koun, E., ···& Farnè, A. (2020). Exploring the effect of cooperation in reducing implicit racial bias and its relationship with dispositional empathy and political attitudes. *Frontiers in Psychology*, 11, Article 510787. https://doi.org/10.3389/fpsyg.2020.510787

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport' s cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, 5(4), 461-476.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751-783.

Phills, C. E., Hahn, A., & Gawronski, B. (2020). The bidirectional causal relation between implicit stereotypes and implicit prejudice. Personality and Social Psychology Bulletin*, 46(9), 1318-1332.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and blacks with approach behaviors. Journal of Personality and Social Psychology*, 100(2), 197-210.

Phills, C. E., Santelli, A. G., Kawakami, K., Struthers, C. W., & Higgins, E. T. (2011). Reducing implicit prejudice: Matching approach/avoidance strategies to contextual valence and regulatory focus. Journal of Experimental Social Psychology*, 47(5), 968-977.

Phills, C. E., Kawakami, K., Krusemark, D. R., & Nguyen, J. (2019). Does reducing implicit prejudice increase out-group identification? The downstream consequences of evaluative training on associations between the self and racial categories. Social Psychological and Personality Science*, 10(1), 26-34.

Qian, M., Heyman, G. D., Wu, M., & Fu, G. (2022). Individuating multiple (not one) persons reduces implicit racial bias. *Frontiers in Psychology*, 13, Article 939811. https://doi.org/10.3389/fpsyg.2022.939811

Qian, M. K., Quinn, P. C., Heyman, G. D., Pascalis, O., Fu, G., & Lee, K. (2017). Perceptual individuation training (but not mere exposure) reduces implicit racial bias in preschool children. *Developmental Psychology*, 53(5), 845-859.

Qiu, Y., & Zhang, G. (2020). Make exercise easier: A brief intervention to influence implicit attitudes towards exercise and physical activity behavior. Learning and Motivation*, 72, Article 101660. https://doi.org/10.1016/j.lmot.2020.101660

Rauf, I., Hartmann, A., Koumtchev, A., Khan, S. A., & Kashyap, R. (2022). Conscious and unconscious bias: The hidden pandemic of biases in healthcare exacerbated by COVID-19. *HCA Healthcare Journal of Medicine*, 3(3), 89-94.

Rees, H. R., Rivers, A. M., & Sherman, J. W. (2019). Implementation intentions reduce implicit stereotype activation and application. Personality and Social Psychology Bulletin*, 45(1), 37-49.

Romano, L. E. (2024). Mindful discipline: A pilot study exploring mindfulness and the reduction of punitive discipline for Black male students. School Psychology*. https://doi.org/10.1037/spq0000654

Sadler, M., & Devos, T. (2020). Ethnic diversity matters: Putting implicit associations between weapons and ethnicity in context. *Group Processes & Intergroup Relations*, 23(2), 285-303.

Stell, A. J., & Farsides, T. (2016). Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions. Motivation and Emotion*, 40(1), 140-147.

Strick, M., & Volbeda, M. (2018). When the valence of unconditioned stimuli evolves over time: Evaluative conditioning with good-ending and bad-ending stories. Journal of Experimental Social Psychology*, 74, 50-55.

Strick, M., Van Baaren, R. B., Holland, R. W., & Van Knippenberg, A. (2009). Humor in advertisements enhances product liking by mere association. Journal of Experimental Psychology: Applied*, 15(1), 35-45.

Shaw, J. A., Forman, E. M., Espel, H. M., Butryn, M. L., Herbert, J. D., Lowe, M. R., & Nederkoorn, C. (2016). Can evaluative conditioning decrease soft drink consumption? Appetite*, 105, 60-66.

Sheeran, P., Listrom, O., & Gollwitzer, P. M. (2025). The when and how of planning: Meta-analysis of the scope and components of implementation intentions in 642 tests. *European Review of Social Psychology*, 36(1), 162-204.

Shen, J., Clinton, A. J., Penka, J., Gregory, M. E., Sova, L., Pfeil, S., ···& Maa, T. (2024). Smartphone-based virtual and augmented reality implicit association training (VARIAT) for reducing implicit biases toward patients among health care providers: App development and pilot testing. *JMIR Serious Games*, 12, Article e51310. https://doi.org/10.2196/51310

Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. Psychological Science*, 19(7), 717-723.

Skorinko, J. L. M., DiGiovanni, C., Rondina, K., Tavares, A., Spinney, J., Kobeissi, M., ···& Doyle, A. (2023). The effects of perspective taking primes on the social tuning of explicit and implicit views toward gender and race. *Frontiers in Psychology*, 14, Article 1014803. https://doi.org/10.3389/fpsyg.2023.1014803

Smith, C. T., Calanchini, J., Hughes, S., Van Dessel, P., & De Houwer, J. (2020). The impact of instruction-and experience-based evaluative learning on IAT performance: A Quad model perspective. *Cognition and Emotion*, 34(1), 21-33.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), 108-131.

Stathi, S., Cameron, L., Hartley, B., & Bradford, S. (2014). Imagined contact as a prejudice-reduction intervention in schools: The underlying role of similarity and attitudes. *Journal of Applied Social Psychology*, 44(8), 536-546.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. Personality and Social Psychology Bulletin*, 34(10), 1337-1347.

Stewart, T. L., Latu, I. M., Kawakami, K., & Myers, A. C. (2010). Consider the situation: Reducing automatic stereotyping through situational attribution training. *Journal of Experimental Social Psychology*, 46(1), 221-225.

Stewart, T. L., Latu, I. M., Martin, T., Walsh, S. P., Schmidt, A., & Kawakami, K. (2022). Implicit bias reduction that lasts: Putting situational attribution training to the test. *Journal of Applied Social Psychology*, 52(11), 1062-1074.

Stone, J., Moskowitz, G. B., Zestcott, C. A., & Wolsiefer, K. J. (2020). Testing active learning workshops for reducing implicit stereotyping of Hispanics by

majority and minority group medical students. *Stigma and Health*, 5(1), 94-102.

Sullivan-Bissett, E. (2023). Virtually imagining our biases. *Philosophical Psychology*, 36(4), 860-878.

Tassinari, M., Aulbach, M. B., Harjunen, V. J., Cocco, V. M., Vezzali, L., & Jasinskaja-Lahti, I. (2024). The effects of positive and negative intergroup contact in virtual reality on outgroup attitudes: Testing the contact hypothesis and its mediators. Group Processes & Intergroup Relations*, 27(8), 1773-1790.

Tassinari, M., Aulbach, M. B., & Jasinskaja-Lahti, I. (2022). Investigating the influence of intergroup contact in virtual reality on empathy: An exploratory study using AltspaceVR. *Frontiers in Psychology*, 12, Article 815497. https://doi.org/10.3389/fpsyg.2021.815497

Tello, N., Jaafari, N., & Chatard, A. (2020). Effects of evaluative conditioning on implicit evaluation of alcohol and drinking behaviors: A direct replication. *Alcohol and Alcoholism*, 55(3), 299-305.

Terbeck, S., Kahane, G., McTavish, S., McCutcheon, R., Hewstone, M., Savulescu, J., ···& Norbury, R. (2015). β-Adrenoceptor blockade modulates fusiform gyrus activity to black versus white faces. Psychopharmacology*, 232(16), 2951-2958.

Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P. J., & Hewstone, M. (2012). Propranolol reduces implicit negative racial bias. Psychopharmacology*, 222(3), 419-424.

Trucharte, A., Carmen, V., Pacios, J., Bruña, R., Espinosa, R., Peinado, V., ···& Bentall, R. P. (2024). Could an evaluative conditioning intervention ameliorate paranoid beliefs? Self-reported and neurophysiological evidence from a brief intervention focused on improving self-esteem. Frontiers in Psychiatry*, 15, Article 1472332. https://doi.org/10.3389/fpsyt.2024.1472332

Turner, R. N., & Crisp, R. J. (2010). Imagining intergroup contact reduces implicit prejudice. British Journal of Social Psychology*, 49(1), 129-142.

Van Dessel, P., De Houwer, J., Gast, A., Roets, A., & Smith, C. T. (2020). On the effectiveness of approach-avoidance instructions and training for changing evaluations of social groups. Journal of Personality and Social Psychology*, 119(2), E1-E14.

Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not explicit evaluation. Journal of Experimental Social Psychology*, 63, 1-10.

Vitriol, J. A., & Banaji, M. R. (2024). Eliciting cognitive consistency increases acceptance of implicit bias. *Journal of Experimental Psychology: General*, 153, 2859-2872.

Vuletich, H. A., & Payne, B. K. (2019). Stability and change in implicit bias. *Psychological Science*, 30(6), 854-871.

Wang, Y., & Zhang, M. (2020, November). Reducing implicit gender biases in software development: Does intergroup contact theory work? In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 580-592).

Webb, T. L., Sheeran, P., & Pepper, J. (2012). Gaining control over responses to implicit attitude tests: Implementation intentions engender fast responses on attitude-incongruent trials. British Journal of Social Psychology*, 51(1), 13-32.

West, K., Hotchin, V., & Wood, C. (2017). Imagined contact can be more effective for participants with stronger initial prejudices. *Journal of Applied Social Psychology*, 47(5), 282-293.

Williams, E. C., & Polito, V. (2022). Meditation in the workplace: Does mindfulness reduce bias and increase organisational citizenship behaviours? Frontiers in Psychology*, 13, Article 747983. https://doi.org/10.3389/fpsyg.2022.747983

Williamson, S., & Foley, M. (2018). Unconscious bias training: The 'silver bullet' for gender equity? *Australian Journal of Public Administration*, 77(3), 355-366.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107(1), 101-126.

Wittlin, N. M., Dovidio, J. F., Burke, S. E., Przedworski, J. M., Herrin, J., Dyrbye, L., ···& van Ryn, M. (2019). Contact and role modeling predict bias against lesbian and gay individuals among early-career physicians: A longitudinal study. Social Science & Medicine*, 238, Article 112422. https://doi.org/10.1016/j.socscimed.2019.112422

Xia, T., Yao, Z., Guo, X., Liu, J., Chen, D., Liu, Q., ···& Hu, X. (2023). Updating memories of unwanted emotions during human sleep. *Current Biology*, 33(2), 309-317.

Zerhouni, O., Houben, K., El Methni, J., Rutte, N., Werkman, E., & Wiers, R. W. (2019). I didn' t feel like drinking, but I guess why: Evaluative conditioning changes on explicit attitudes toward alcohol and healthy foods depends on contingency awareness. Learning and Motivation*, 66, 1-11.

Zhu, X., & Liao, Q. (2020). Research on implicit occupational gender stereotype based on IAT. *Advances in Psychology*, 10(02), 124-135.

Zhang, Q., Zhang, G., & Dang, J. (2023). Enjoy exercise more: The effects of dual-channel evaluative conditioning on implicit attitudes toward exercise. Scandinavian Journal of Medicine & Science in Sports*, 33(11), 2381-2389.

Note: Figure translations are in progress. See original paper for figures.