

内隐社会认知的干预方法*

郭小丽 常骏垚 沙吗阿杰 杨紫嫣

(中国科学院心理研究所, 认知科学与心理健康全国重点实验室, 北京 100101)
(中国科学院大学心理学系, 北京 100049)

摘要 自 1995 年内隐社会认知概念提出以来, 该领域吸引了大量研究, 涵盖内隐态度、内隐刻板印象和内隐自尊等多个方面。早期研究主要集中于内隐社会认知的测量及其影响。研究表明负面的内隐社会认知(如内隐偏见、内隐刻板印象、低内隐自尊)对人们的心理和行为具有一系列消极影响。近年来, 研究焦点逐步转向如何有效干预负面的内隐社会认知, 以期减少其不良后果。目前常用的有效干预方法包括评价性条件反射、接近-回避训练、群际接触、接触反刻板印象案例、内隐偏见培训、实施意图、靶向记忆再激活、诱导情绪和正念冥想 9 种策略。本文对上述干预方法及其效果进行了系统介绍, 并对每种方法的效应量进行了元分析。结果表明, 这些方法表现出小到中等程度的效应量, 其中接近-回避训练、内隐偏见培训和情绪诱导干预较为有效。上述干预方法在短期内表现出一定的有效性, 但缺乏长久的时效性。后续研究可以采用多方法结合提高干预效果及时效性。此外, 未来的内隐社会认知干预可以在人工智能技术的支持下更加高效和便捷。

关键词 内隐社会认知, 内隐联系测验, 内隐偏见, 内隐刻板印象, 干预策略

分类号 00000

1 引言

随着社会认知研究的发展, 人们逐渐意识到, 许多心理和行为并非完全由显性意识驱动, 而是深植于个体的无意识层面(Nisbett & Wilson, 1977)。20 世纪 90 年代, “内隐社会认知”概念的提出, 为理解人类行为的潜在心理机制提供了新的视角。根据内隐社会认知理论, 人们的社会心理和行为常常受到无意识的、自动化心理过程的影响。内隐社会认知逐渐成为社会心理学领域的焦点, 其内容聚焦于内隐态度、内隐刻板印象、内隐自尊等方面。大量研究揭示内隐社会认知对个体心理和行为具有重要的作用, 甚至对宏观社会的教育、就业和司法系统具有广泛的影响(杨紫嫣等, 2015; Greenwald, Dasgupta, et al., 2022; Greenwald & Lai, 2020)。

随着对内隐社会认知的深入理解, 尤其是负

面的内隐社会认知(如内隐偏见、内隐刻板印象、低内隐自尊等)可能导致的不良影响, 例如偏见、歧视、社会不公等(杨紫嫣等, 2015; Greenwald & Lai, 2020), 研究者们越来越关注如何有效地干预这些内隐社会认知, 以期降低其不良后果。为此, 研究者们开发了一系列内隐社会认知的干预方法, 包括评价性条件反射、接近-回避训练、群际接触、接触反刻板印象案例等等, 并通过实证研究验证了这些方法的有效性。本文将对有效的内隐社会认知干预方法进行系统介绍, 并通过元分析揭示各类方法的干预效果。通过对现有干预方法的整合, 本文希望为未来的内隐社会认知干预的研究和实践提供依据, 并探讨未来改善干预效果的可能方向。

2 内隐社会认知的内涵和测量

内隐社会认知主要描述的是自动化、无意识的心理过程, 这些过程在个体未察觉的情况下影响其心理和行为(Greenwald & Banaji, 1995)。内隐社会认知主要包括内隐态度、内隐刻板印象和内隐自尊等。内隐态度是指个体在无意识层面所持

收稿日期: 2025-03-16

* 国家自然科学基金面上项目(32371130)。

通信作者: 杨紫嫣, E-mail: yangzy@psych.ac.cn

有的、自动化的评价性反应或情感倾向,它通常对特定的对象、群体产生影响(Greenwald et al., 2002; Wilson et al., 2000)。内隐偏见通常被视作是内隐态度的特殊表现形式或负面的内隐态度(Gawronski, 2024; Lai et al., 2013)。研究表明,内隐偏见广泛存在,个体往往在不自觉的情况下表现出偏见行为,且通常未能意识到自己的行为受到偏见的影响(Gawronski et al., 2022; Greenwald, Dasgupta, et al., 2022)。内隐刻板印象指个体无意识地对特定群体所持的自动化、概括性的认知或信念(Greenwald & Banaji, 1995)。内隐刻板印象广泛存在于性别、种族、年龄、职业等社会类别中,在个体不知情的情况下影响其对群体成员的判断和行为。内隐自尊是个体自动化、无意识的自我情感性评价(Falk & Heine, 2015; Greenwald et al., 2002),内隐自尊对个体的心理健康具有重要的作用(杨紫嫣 等, 2015)。

内隐社会认知通常采用间接测量方法,其中最常用的是内隐联系测验(Implicit Association Test, IAT) (Greenwald, Brendl, et al., 2022)。IAT采用计算机分类任务,通过相容任务(如“白人-积极”)和不相容任务(如“黑人-消极”)的反应时之差来衡量概念与属性之间的自动联结,从而间接测量个体的内隐态度、刻板印象、自尊等心理(Greenwald et al., 1998)。IAT作为内隐测量中最普及和有效的工具,在内隐社会认知研究中广泛应用(Cai & Wu, 2021; Jin et al., 2022; Zhu & Liao, 2020)。

3 内隐社会认知的稳定性和可变性

内隐社会认知是稳定的还是可变的一直是社会心理学领域备受关注且存在争议的问题。早期的慢学习过程理论认为,个体在成长过程中通过渐进式学习逐步形成内隐偏见等内隐心理,因此内隐社会认知会随着个体的自然发展而发生变化(Smith & DeCoster, 2000)。与此相对,类别学习理论提出,个体一旦掌握了某一社会类别的概念,所有与该类别相关的联结会自动形成并保持稳定,因此内隐社会认知在个体的发展过程中应当是稳定不变的(Baron, 2015; Dunham et al., 2008, 2013)。

近年来,越来越多的研究发现,不论个体层面还是群体层面的内隐社会认知都是可变的(Marini

et al., 2024; Vuletich & Payne, 2019)。内隐社会认知的变化受短期情境因素和长期文化因素的共同作用。例如,特定的社会信息(启发多元文化)能够在短期内引发内隐态度变化,而持久的态度转变则需长期的教育、政策和文化共同影响(Lai et al., 2013)。从个体来看,Cvencek等人发现内隐刻板印象(如性别-学科)在个体发展过程中呈现出加深的趋势,特别是在学龄儿童和青少年群体中,个体的内隐刻板印象随着年龄的增长逐渐稳固(Cvencek et al., 2014, 2015)。Cai等人通过追踪研究进一步揭示了内隐社会认知既具有稳定性,也存在随时间变化的潜力。具体而言,内隐自尊稳定性较低,整体呈下降趋势;性别-学科相关的内隐刻板印象则展现出较强的稳定性;而内隐种族态度的稳定性则处于中等水平,其个体间的变化表现出较高的异质性(Cai et al., 2021)。从群体来看,Charlesworth等人的一系列研究揭示了外显偏见总体上呈现减少趋势,而内隐偏见的变化模式较为复杂,受到个人发展和社会文化环境的影响(Charlesworth & Banaji, 2019, 2022a, 2021)。其中,内隐种族偏见在长期文化和教育干预下更易改变,而性别偏见更稳定(Charlesworth & Banaji, 2022b)。认知神经研究揭示了内隐社会认知变化的生理机制,前额叶皮层和杏仁核等脑区活动与内隐偏见的改变密切相关(高可翔 等, 2023; Cai & Wu, 2021)。

综上所述,内隐社会认知既具有一定的稳定性,又会因情境因素而表现出可变性。鉴于此,研究者们致力于探索多种干预策略,以期降低负面的内隐社会认知(如内隐偏见),提升积极的内隐社会认知(如内隐自尊)。下文将系统介绍当前主要的内隐社会认知干预方法。

4 内隐社会认知的干预方法及效果

基于已有研究,本文对主要的内隐社会认知干预方法进行了整理,下文将对各种方法及其效果展开详细介绍。本文对这些方法的干预内容、干预原理、有效性、优缺点等进行了系统总结(表1)。本文进一步对各类方法的效应量进行了元分析,首先将原文研究报告的效应量统一转换为Cohen's *d*,之后基于各类方法的Cohen's *d*,采用Jamovi 软件 Major 模块进行了元分析(表1)。

表 1 内隐社会认知的干预方法总结

干预方法	干预原理	干预内容	效应量 范围(<i>d</i>)	元分析 效应量(<i>d</i>)	时效性	优点	局限性
评价性条件反射	经典条件反射理论 (Olson & Fazio, 2006)	内隐种族偏见	0.18~1.24	0.44	短期有效	简单易操作, 适用于多种内隐偏见类型	效果短暂, 需重复强化干预
		内隐自尊	0.38~0.83				
		内隐刻板印象	0.36~0.59				
		内隐态度	0.11~2.05				
接近-回避训练	行为主义理论 (Kawakami et al., 2007)	内隐种族偏见	0.11~1.02	0.53	短期有效	配合肢体动作, 参与感更强	依赖实验室环境, 限制了现实推广性
		内隐学科态度	1.17				
群际接触	群际互动理论 (Pettigrew & Tropp, 2006)	内隐种族偏见	0.02~1.17	0.45	短期有效	促进实际互动, 便于实践	接触机会有限, 但可以使用VR替代
		内隐群体偏见	0.32~1.34				
接触反刻板印象案例	认知重构理论 (Lai et al., 2014)	内隐种族偏见	0.03~0.66	0.32	短期有效, 对年长儿童和成人效果更佳	创新性较强, 适合教育场景	干预效果不稳定
		内隐年龄偏见	0.9				
内隐情境归因培训	认知行为主义理论 (Devine et al., 2022)	内隐种族偏见	0.46~1.67	0.58	具有长期有效性潜力	综合性较强, 适用于多种情境	效果不稳定、会因培训设计和文化背景而异, 有时甚至起反作用
		内隐种族偏见	0.39~0.56				
		内隐刻板印象	0.34~1.74				
		内隐种族偏见	0.23~0.44				
实施意图	目标设定理论 (Gollwitzer et al., 2006)	内隐种族偏见	0.38	0.34	短期有效	简单易于操作	效果短暂, 需结合其他方法
		内隐刻板印象	0.23~0.81				
靶向记忆再激活	认知重构+记忆巩固理论 (Hu et al., 2015)	内隐种族偏见	0.04~0.85	0.49	具有长期有效性潜力	理论创新性较强, 结合睡眠巩固干预记忆	技术要求高, 需进一步验证稳健性
		内隐自尊	0.46~0.69				
诱导情绪	认知重构+情绪调节理论(Huntsinger et al., 2009)	内隐种族偏见	0.46~0.8	0.55	短期有效	灵活性高, 适用于多种情境、可以和多种方法结合使用	干预时效性有限、需和其他方法结合使用
		内隐刻板印象	0.54~1.04				
		内隐态度	0.17~0.72				
		内隐自尊	0.17~0.51				
正念冥想	认知重构+正念理论 (Kang et al., 2014)	内隐种族偏见	0.15~0.94	0.34	具有长期有效性潜力	适用于多种情境, 提升整体心理健康	需要结合其他干预方法并且长期练习
		内隐态度	0.02~0.56				
		内隐刻板印象	0.15~0.92				

4.1 评价性条件反射(Evaluative Conditioning, EC)

评价性条件反射是一种基于巴普洛夫经典条件反射理论的程序, 广泛应用于态度的形成与改变。其核心策略是通过反复将条件刺激(Conditional Stimulus, CS)与具有明显情感效价的

无条件刺激(Unconditional Stimulus, US)配对, 从而促使个体对条件刺激的态度或评价发生变化(Moran et al., 2023)。大量实证研究验证了EC在内隐偏见干预中的有效性。例如, Olson等通过将黑人面孔(CS)与积极词汇(US)进行配对, 发现参

与者对黑人群体的内隐态度变得更加积极,表明EC能够降低内隐种族偏见(Olson & Fazio, 2006)。类似地,将正面情感与特定群体(如黑人群体)建立联结,可以有效降低对该群体的内隐偏见,并增强个体与该群体的正向情感联结(Phills et al., 2019),改善群体间的内隐态度,进一步推动群体态度转变(Kurdi et al., 2019)。另有研究进一步表明,EC能够减少对特定种族的内隐偏见,如对穆斯林和阿拉伯群体的偏见(French et al., 2013)。Lai等人通过系列研究验证了EC对于减少内隐种族偏见的有效性,但发现该干预的时效性有限,均未超过24小时(Calanchini et al., 2021; Lai et al., 2014; Lai et al., 2016)。

EC的应用不限于内隐偏见,其在减少内隐偏见的同时还能有效降低内隐刻板印象(Phills et al., 2020)。如使用基于EC逻辑的“反刻板印象训练”来减少内隐刻板印象,通过呈现一系列与刻板印象不符的刺激组合,例如将“女性”这一社会类别和“强壮”等特质词汇相关联并要求参与者对这些非刻板组合做出反应以形成反刻板印象联想。结果表明,反刻板训练成功减少了内隐性别、种族刻板印象(Gawronski et al., 2008)。

EC还能够增强个体特定的内隐态度。例如, Qiu和Zhang发现,EC能够改善人们对锻炼的内隐态度,尽管其对实际锻炼行为的影响较为有限(Qiu & Zhang, 2020; Zhang et al., 2023)。Jeon等(2019)的研究表明,评价性条件反射(Evaluative Conditioning, EC)能有效改善青少年对暴力行为的内隐态度。此外,EC还被证明可以改变个体对饮酒(Zerhouni et al., 2019)、饮食(Shaw et al., 2016)以及特定品牌(Charlesworth et al., 2020)的内隐态度。

此外,还能够应用EC来增强内隐自尊。多项研究表明,通过将“我”这一自我概念与积极的特质词汇进行配对,能够有效提升个体的内隐自尊(Baccus et al., 2004; Dijksterhuis, 2004; Grumm et al., 2009)。最新的研究表明,EC对内隐自尊的增强效果不仅在正常人群中有效,亦适用于临床群体(如偏执症患者)(Trucharte et al., 2024)。综上,EC通过将正面或负面情感刺激与特定目标对象关联,能够有效改变内隐社会认知,并进一步影响个体的行为(Luo et al., 2024; Tello et al., 2020)。

4.2 接近-回避训练(Approach-Avoidance Training)

接近-回避训练(Approach-Avoidance Training,

AAT)是一种通过身体接近或回避特定刺激来影响个体态度的方法。通过训练参与者接近黑人(看到黑人面孔图片时使用操纵杆将图片拉向自己)避开白人(看到白人面孔图片时使用操纵杆将图片推离自己),随后用种族IAT测量种族偏见。结果发现,重复接近某一社会群体能够改善对该群体的内隐态度,并且这种训练在实际跨种族互动中促进了非言语行为(如微笑、眼神接触)(Kawakami et al., 2007)。AAT在降低内隐种族偏见方面的有效性得到了后续研究的支持(Van Dessel et al., 2016, 2020)。Phills等人进一步扩展了AAT的应用,研究了策略(接近或回避)和情境(正面或负面)的匹配对干预效果的影响。结果表明,在正面情境中,接近策略显著减少内隐偏见,而在负面情境中,回避策略更为有效,说明了策略与情境匹配的重要性(Phills et al., 2011)。

在AAT过程中增加特定指示,即接近-回避指令(AA指令),也能够在内隐偏见干预中发挥一定的效果。AA指令指在实验或训练过程中,指导参与者对有关刺激进行接近或回避的特定操作指示,以确保参与者按照预定的规则进行操作。近期研究发现,相较于传统的AAT,增加AA指令对于改变个体对不熟悉食品品牌的内隐评价具有更大的效应(Smith et al., 2020),进一步拓展了AAT在特定领域内隐态度干预中的潜力。综上,接近-回避训练通过引导个体对特定刺激进行接近或回避反应,关键时刻辅以指令能够有效地干预内隐偏见及更广泛的内隐态度。

4.3 群际接触(Intergroup Contact)

群际接触理论作为减少群体间偏见的重要策略,已扩展至内隐偏见干预领域(Paluck et al., 2019; Pettigrew & Tropp, 2006)。群际接触可分为直接接触和间接接触。大量研究表明,直接和间接接触均能有效减少群体间的内隐偏见(Bond et al., 2023)。就直接接触而言,Shook和Fazio(2008)通过大学宿舍系统的自然实验,将白人新生随机分配至与非裔美国人或白人室友同住,发现那些与不同种族室友同住的学生,其内隐种族态度和偏见表现出显著改善,而与同种族室友同住的学生未见明显变化。同样,Wang通过将软件开发人员分配至不同男女比例的研究开发团队中,发现不同性别的直接接触能够减少对软件开发职业的内隐性别偏见(Wang & Zhang, 2020)。另有研究揭示

了接触质量和频率对干预效果具有显著影响(Farmer et al., 2020; Onyeador et al., 2020; Wittlin et al., 2019)。最近一项纵向研究表明, 高质量且频繁的接触能够显著减少内隐偏见, 并预测外群体内隐态度的改善, 为群际和谐提供了实践指导(Di Bernardo et al., 2022)。此外, 在实施群际接触干预策略时, 应鼓励积极的互动方式, 并尽量减少消极的接触体验(Barlow et al., 2012)。

直接接触可以有效干预内隐偏见, 针对无法实现直接接触的情境, 研究者开发了间接接触(如想象群际接触、虚拟现实接触)策略。想象群际接触能有效减少对特定群体的内隐偏见(Stathi et al., 2014; Turner & Crisp, 2010; West et al., 2017)。如通过想象群际接触减少对农民工、同性恋及肥胖人群的内隐偏见, 值得注意的是, 接触类型及个体先前的接触经验会影响干预效果。随着虚拟现实(Virtual Reality, VR)技术的发展, 研究人员尝试将群际接触拓展至虚拟现实层面。Tassinari 等研究了 VR 中的群际接触对内隐态度的影响。参与者通过 VR 沉浸在虚拟环境中, 以白人身份与黑人(群际接触)或白人(群内接触)玩家进行互动。研究发现, 通过 VR 与黑人的合作性接触可以改善对黑人的内隐态度(Tassinari et al., 2024)。研究结果支持了群际接触理论, 并将其应用范围扩展至虚拟现实环境(Doan et al., 2021; Sullivan-Bissett, 2023), 模拟现实世界中难以实现的积极群际互动从而降低内隐偏见。综上, 直接和间接的群际接触已广泛应用于减少群体间的内隐偏见, 且群际接触的干预效果可延伸至内隐刻板印象领域(Eisenstadt et al., 2024)。具体而言, 通过增强群体间的正面互动能够增加对移民的积极情绪, 降低对移民的负面内隐刻板印象, 减轻因移民而感知到的威胁(Duque et al., 2024)。

4.4 接触反刻板印象案例(Exposure to Counter-stereotypical Exemplars)

接触反刻板印象案例是指通过接触与刻板印象不符的特定形象, 从而减少对特定群体的内隐偏见和刻板印象, 具体包括以下 4 种干预方法。首先, 生动反刻板场景(Vivid Counter-Stereotypic Scenario)是一种通过情感和情境沉浸接触反刻板印象场景的干预方法。在生动的反刻板场景中, 参与者以第一人称视角阅读反刻板故事(例如, 故事中黑人是正面形象, 白人是负面形象, 并且黑

人英雄还拯救了被白人歹徒袭击的“我”)。这种方法能减少对黑人的内隐种族偏见(Marini et al., 2011)。第二, 反刻板印象 IAT 练习(Practicing An IAT With Counter-Stereotypical Exemplars)将 IAT 和反刻板印象典型案例相结合。Lai 等(2014)在实验中使用正面的典型黑人和负面的典型白人案例来增强对黑人的积极联结和对白人的消极联结, 有效减少了参与者对黑人的内隐种族偏见。第三, 通过竞争改变群体边界(Shifting Group Boundaries Through Competition)旨在通过模拟团队合作和竞争来改变对不同种族群体的认知。在模拟的躲避球游戏中, 参与者与黑人队友合作对抗白人对手, 其中黑人队友展现出积极行为, 而白人对手则表现出负面行为。研究表明, 这种方法能有效减少对黑人群体的内隐偏见(Lai et al., 2014)。第四, 威胁情境群体归属(Shifting Group Affiliations Under Threat)是一种通过极端威胁情境干预内隐态度的方法。参与者阅读一个关于核战后威胁场景的故事(其中黑人角色被描绘为具有积极生存技能的“亲密朋友”, 而白人角色则是企图破坏营地的“可怕敌人”), 结果表明这样的方式能够减少参与者对黑人的负面偏见(Park et al., 2007)。接触反刻板印象案例的干预方法同样适用于减少儿童的内隐种族偏见, 但是其干预效果在较年幼的儿童(5~8岁)中不如年长儿童(9~12岁)(Gonzalez et al., 2017, 2021)。其他研究表明, 有时仅依赖接触反刻板印象的榜样并不足以有效减少内隐偏见, 需要设计更具个性化的训练方案, 以针对性地降低内隐偏见(Qian et al., 2017; Qian et al., 2022)。

综上, 反刻板印象案例在减少内隐种族偏见方面表现出一定的有效性。然而, 现有的干预方法主要在实验环境中验证, 在真实世界中的效果和持久性尚未得到充分评估, 也有研究指出这些方法缺乏持久效应(FitzGerald et al., 2019; Lai et al., 2016), 且单一的接触反刻板印象案例有时候难以见效, 需要同其他方法相结合。

4.5 内隐偏见培训(Unconscious Bias Training, UBT)

内隐偏见培训, 或称内隐偏见导向的多元化培训(Implicit Bias-Oriented Diversity Training), 通过结合多种认知干预策略提升个体对内隐偏见的认知与控制, 进而减少内隐偏见(Hagiwara et al., 2024; Paluck et al., 2021)。UBT 内容包含教育

和培训两个核心部分: 教育部分向参与者传达内隐偏见的概念, 帮助他们理解内隐偏见的存在及其对个人决策和社会行为的影响, 特别是潜在的负面后果; 培训部分包括若干干预策略, 旨在通过这些干预措施促使参与者从不同视角重新审视外群体, 减少对特定群体的内隐偏见。UBT 应用范围广泛, 并且致力于将干预应用于实际(Devine & Ash, 2022; Möller et al., 2024)。下面展开介绍几种 UBT 中常用的干预措施。

4.5.1 情境归因训练(Situational Attribution Training, SAT)

情境归因训练通过引导人们更加注重外部情境因素, 减少将外群体成员的负面行为归因于其内在特质的倾向, 即基本归因错误(Pettigrew, 1979)。例如, Stewart 等人发现接受 SAT 的白人参与者更倾向于使用情境因素来解释黑人男性的负面行为, 表现出更少的内隐种族刻板印象(Stewart et al., 2010), 并且情境性归因能够减少群体间的内隐偏见以及后续负面行为(Levontin et al., 2013)。Stewart 等人最近的研究通过更加密集的训练设计, 增加了训练的深度和广度, 发现 SAT 的效果在 24 小时后仍然显著, 并且这种效果可以推广到训练中不涉及的负面内隐刻板印象。近期的研究显示了 SAT 的时效性潜力, 且表明通过重复训练或结合其他干预措施, 或能进一步延长其干预效果(Stewart et al., 2022)。总的来说, SAT 的持久性和广泛适用性使其成为内隐偏见干预的重要方法。

4.5.2 偏见习惯打破(Prejudice Habit-Breaking Intervention, PHBI)

偏见习惯打破干预理论将内隐偏见和刻板印象视为一种“习惯”, 认为可以通过自主识别、调节以及系统行为进行有效干预, 其内涵同 UBT 一致均包括教育和培训两个部分(Devine et al., 2012; Forscher & Devine, 2014)。研究证明, PHBI 是减少内隐偏见的有效方法之一, 并且具有长时效应。例如, Devine 等人(2012)进行的一项为期 3 个月的纵向研究发现, 接受 PHBI 干预的实验组的内隐种族偏见显著下降, 同时对歧视问题的关注显著增强, 并且这种效果直至第 8 周仍然存在, 而对照组未表现出类似的变化(Devine et al., 2012)。PHBI 在实际社会情境中也发挥作用(Cox & Devine, 2019)。例如, 多个研究显示, PHBI 在

STEM (科学、技术、工程和数学)领域的干预减少了内隐性别偏见和内隐性别刻板印象, 显著提高了女性招聘比例, 促进性别多样性和社会公平(Carnes et al., 2015, 2023; Devine et al., 2017)。在最新的研究中, PHBI 被应用于瑞典社会工作者的内隐偏见培训。研究结果表明, 尽管干预组内隐偏见的降低并未取得预期效果, 但参与者对于内隐偏见的关注度得到了显著提升(Bursell, 2024)。这一发现说明 PHBI 的效果可能受到特定社会和文化背景的影响, 强调了将实验室中有效的干预措施转化为实际应用时, 须考虑干预措施的普遍性和适用性。

4.5.3 启动多元文化(Priming Multiculturalism)

启动多元文化通过促使个体认识、欣赏种族和文化多样性, 来减少对特定群体的内隐偏见。参与者被引导阅读并反思多元文化主义理念, 随后通过 IAT 测量内隐种族偏见。结果发现, 启动多元文化主义有效减少了内隐种族偏见(FitzGerald et al., 2019; Lai et al., 2014; Lai et al., 2016)。Salder 等人通过研究美国 351 个大都会区域的种族多样性, 探讨了其对内隐偏见(如黑人与武器的内隐关联)的影响。研究发现, 提升种族多样性和种族融合度能有效减弱对黑人的内隐偏见(Sadler & Devos, 2020)。参加多元化培训(接触、交流以及促进社会认同等)的个体的内隐偏见也有所降低(Behm-Morawitz & Villamil, 2019)。这表明通过多样性培训, 推动种族、群体多样性和融合度, 比单一增加少数群体比例更能有效减少内隐偏见。

综上, 作为一种综合性的新兴干预方法, UBT 可以帮助人们提升偏见认知, 指导人们识别并管理偏见, 从而在日常生活中更加有意识地避免产生偏见及其带来的负面影响(Bezyak et al., 2024; Cooper et al., 2022; Howell et al., 2024; Vitriol & Banaji, 2024)。多数研究揭示了 UBT 的有效性(Jackson et al., 2014; Liang et al., 2019; Liu et al., 2022; Stone et al., 2020), 但也有研究发现 UBT 无效甚至起到反作用(Williamson & Foley, 2018)。例如, Lai 的研究考察了内隐偏见的短期多样性培训是否能减少警察的内隐偏见, 结果表明, 尽管培训提升了警察对偏见的意识, 但并未能有效降低他们的内隐偏见(Lai & Lisnek, 2023)。UBT 的时效性也需要通过外界的不断加强和巩固才能保持(Lindvall-Östling, 2024)。

4.6 实施意图(Implementation Intentions)

实施意图是一种以“如果-那么”逻辑结构为基础的策略,通过将预期目标行为与具体情境线索联结,从而实现目标的条件反射形式(Sheeran et al., 2025)。Lai 等(2014)探讨了实施意图对降低内隐种族偏见的作用。在实验中,白人参与者被训练形成特定的实施意图,例如“如果我遇到一个黑人面孔,那么我将立即产生积极反应”。结果表明,这种方法在短期内显著减少了内隐种族偏见,但其长期效果尚未得到验证(Calanchini et al., 2021; Lai et al., 2016)。此外,实施意图在减少内隐刻板印象方面也显示出了有效性(FitzGerald et al., 2019; Rees et al., 2019; Stewart & Payne, 2008)。例如, Mendoza 等(2010)通过两项研究检验了分心抑制和反应促进这两种实施意图对刻板印象任务准确性的影响。结果表明,两种实施意图均能够提高任务表现的准确性,即抑制内隐刻板印象的行为表现,且这种效应并非通过延长任务反应时间实现,表明实施意图是通过反射性控制而非增加反应时间发挥作用(Mendoza et al., 2010)。综上,实施意图通过将目标行为与具体情境线索相结合,形成了一种自动化的条件反射,有助于个体通过反射性控制自动产生积极的反应,从而短期内降低内隐偏见、内隐刻板印象及相应行为。

4.7 靶向记忆再激活(Targeted Memory Reactivation, TMR)

睡眠对强化和巩固清醒时形成的记忆具有重要作用,睡眠过程中自发的记忆再激活,这一过程可以通过靶向记忆再激活(Targeted Memory Reactivation, TMR)技术进行操纵。TMR 通过在睡眠期间呈现与之前学习经验相关的感官线索(如特定的声音或气味)来增强记忆的巩固效果(Carbone & Diekelmann, 2024)。Hu 等人首次将 TMR 应用于内隐偏见的干预,发现接受“反刻板印象训练”,即有意识构建与刻板印象相悖的联想,以形成反刻板印象联想,并随后接受 TMR 的参与者,其内隐偏见显著减少,且效果至少持续了一周(Hu et al., 2015),强调了记忆巩固在长期干预效果中的重要性。然而, Humiston 等的研究未能重复 Hu 等人的结果,他们的研究发现,无论是在午睡后立即测量还是在一周后测量, TMR 对内隐偏见都没有显著影响(Humiston & Wamsley, 2019)。基于此,研

究者强调需要使用更敏感的分析方法来验证 TMR 在此类干预中的长期有效性(Aczel et al., 2015)。针对不一致的研究结果, Hu 团队对数据进行了重新分析,发现在特定的睡眠阶段进行干预,特别是在慢振荡上升阶段实施 TMR, 可以更有效地减少内隐社会偏见,并且这种效果在一周后的测试中仍然持续(Xia et al., 2023)。此外, Chen 等人的最新研究扩展了 TMR 的应用范围,探索了将自尊认知训练与 TMR 相结合,以提升内隐自尊。研究结果显示,通过在睡眠期间使用特定的听觉线索激活记忆,可以长期有效地促进内隐自尊的改善(Chen et al., 2023)。综上, TMR 作为一种有效的内隐社会认知干预策略且具有长期时效性,但仍需更多研究进一步证明干预效果的稳健性。

4.8 诱导情绪(Emotion Inducing)

内隐社会认知很大程度受到情绪的影响。研究表明,情绪状态能够调节个体内隐刻板印象的激活与表达,具体而言,消极情绪相比于积极情绪,能够减少内隐刻板印象和态度的表达(Huntsinger et al., 2009)。然而,在特定情境下,例如个体持有反刻板印象思维或平等主义态度时,积极情绪也能够有效减少负面内隐社会认知的激活(Huntsinger et al., 2010)。另有研究表明,情绪能力通过影响评价性条件反射,进而调节内隐态度的形成与变化(Hasford et al., 2018)。但是,诱导诸如愧疚感或道德感的情绪难以有效降低个体的内隐种族偏见(Lai et al., 2016; Lai et al., 2014)。

情绪诱导干预具有相应神经基础,例如,白人在看到黑人面孔时杏仁核激活更显著(Lieberman et al., 2005)。基于此,研究者基于药理学的神经干预探讨了药物干预治疗对内隐偏见的作用。例如,通过药物可以降低杏仁核对情绪刺激的反应强度(Hurlemann et al., 2010),据此, Terbeck 发现使用普萘洛尔(Propranolol, 一种非选择性 β -受体阻滞剂)这类药物治疗干预能够显著降低内隐种族偏见(相比于安慰剂组)(Terbeck et al., 2012, 2015)。

4.9 正念冥想(Mindfulness Meditation)

正念冥想是一种通过培养“正念”状态来增强个体对当下经验的觉察与接受的冥想实践。在这一过程中,个体将注意力集中于当前的体验,并以非评判、开放且好奇的态度去观察自身的感受、思维与情绪。这种冥想方式已广泛应用于心理学多个领域,尤其在促进心理健康、减轻压力及改

善情绪调节等方面展现出显著的效果(Ngan & Cheng, 2022)。研究发现正念冥想在改变内隐社会认知方面具有一定的作用,例如减少内隐刻板印象和群体偏见,提升个体的内隐自尊(Djikic et al., 2008; Kang et al., 2014; Narayan, 2019)。正念冥想还可以减少自动化的偏见性思维,进而促进亲社会行为。Lueke 等发现简短的正念冥想音频干预能够减少内隐种族和性别偏见(Hunsinger et al., 2019; Lueke & Gibson, 2015)。

正念冥想的应用范围广泛,涉及教育、职场及临床等情境。近期的研究表明,在处理涉及黑人学生的教育情境时,教师的正念水平调节了对学生的内隐偏见。正念水平较高的教师在对待黑人学生时表现出较少的种族偏见,而正念水平较低的教师在处理涉及黑人学生的问题时可能表现出更多的偏见(Ash et al., 2023; Hirshberg et al., 2022)。这一结果说明通过提升教师正念水平有助于减少偏见性决策,从而减少不公平教育。进一步研究发现,通过正念干预能够提高教师的正念水平,显著降低其对黑人学生的内隐偏见(Romano, 2025)。然而,在职场情境中,通过正念冥想干预来减少内隐年龄偏见的效果并不显著(Williams & Polito, 2022)。这表明单一的正念冥想训练对减少内隐偏见的干预效果尚不稳定。

5 小结与展望

内隐社会认知对个体的心理和行为,甚至宏观社会都具有重要的影响(杨紫嫣 等, 2015; Greenwald & Lai, 2020)。内隐社会认知具有稳定性,也受到情境因素的影响,表现出一定的可变性(Kurdi & Charlesworth, 2023; Rauf et al., 2022)。基于此,近年来研究者致力于探索干预内隐社会认知的方法,以期降低负面内隐社会认知(如内隐偏见、内隐刻板印象、低内隐自尊等)及其导致的消极个人和社会影响。本文对已有的内隐社会认知干预的研究进行了整合和分析,总结出了9种干预方法。根据方法的定义和原理,这些方法可以分为以下4种类别:(1)认知重构类:通过改变个体的认知或联想模式来干预内隐社会认知,如评价性条件反射、接近-回避训练、接触反刻板印象案例、内隐偏见培训、实施意图;(2)社会互动类:通过与个体或群体进行互动来减少偏见,如群际接触;(3)记忆强化类:通过记忆再激活或巩

固来干预内隐社会认知,如靶向记忆再激活;(4)情绪调节类:通过调节个体的情绪或心境状态来干预内隐社会认知,如诱导情绪、正念冥想。

上述方法均表现出一定的有效性,但是其效果在不同的研究中,针对不同的干预内容存在较大的差异。本文对每种方法的效应量(基于Cohen's d)进行了元分析(表1),结果表明:这9种干预方法具有小到中等程度的效应量(0.32~0.58),其中接近-回避训练、内隐偏见培训、诱导情绪干预较为有效,表现出中等程度的效应量。某些方法如实施意图和靶向记忆再激活的研究数量较少,因此元分析结果的可靠性需要未来研究进一步检验。需要注意的是,本文重点分析了实证研究验证有效的干预方法,未对某些单独使用无效的干预方法如观点采择、同理心训练等进行介绍(Calanchini et al., 2021; Skorinko et al., 2023)。研究表明,采用多种方法结合干预可以有效地提高干预效果。例如, Gabrielli 等通过结合群际接触策略以及同理心训练、观点采择,结果发现,接受多策略干预的儿童对移民的内隐偏见显著减少(Gabrielli et al., 2022)。此外,正念冥想结合人际关系促进的方法能够有效减少预备教师的内隐种族偏见,且在6个月的随访中效果仍旧存在(Hirshberg et al., 2022)。

尽管上述干预方法在短期内都取得了一定成效,但很多方法的时效性通常限于短期,尚未表现出显著的长期效应,例如接近-回避训练、群际接触等,甚至有的方法在一定情境下表现出反作用,例如内隐偏见培训和消极群际接触(Williamson & Foley, 2018)。由此可见,干预效果的时效性成为现有干预方法的重要问题,主要原因可以归结为以下三点:第一,大多数干预方法往往依赖实验室实验,忽略了情境因素,限制了方法在现实世界中的推广性;第二,现有干预措施通常过于单一,干预次数较少,未能形成持续强化的效应;第三,现有干预方法主要针对改变认知本身,而忽略了内隐社会认知形成过程中社会文化因素的影响。研究表明,具有强时效性的干预方法通常具有高度情境化的特点,且需要在特定的受众群体中,通过长时、重复的或多方法结合才能达到最佳效果(Chapman et al., 2018; Devine et al., 2017; Stone et al., 2020)。因此,未来的干预研究可以注重以下三个方向:一是开发更具针对性和情境性的干预策略,以确保其生态效度;二是采用多方

法结合并进行重复强化干预,以提高长期效果;三是探讨文化和社会背景对内隐社会认知的影响,明确干预方法的适用边界,从而确保干预的适用性。

随着人工智能技术的发展,内隐社会认知的干预方法可以从传统的实验室环境扩展到更生态的应用场景。例如,基于虚拟现实(Virtual Reality, VR)、自然语言处理等的干预方法,可以推动未来内隐社会认知的干预。VR能够模拟出逼真、沉浸的社会场景,通过VR个体可以体验不同的社会角色和身份,感受多样化的社会互动和文化差异(Marini & Casile, 2022; Patané et al., 2020)。VR干预还可以通过个性化设计,根据个体的内隐偏见特征,针对性地设置情境,强化沉浸式干预效果(Shen et al., 2024)。这种高度沉浸、个性化的体验可以有效干预内隐社会认知(Tassinari et al., 2022, 2024)。VR技术在识别内隐偏见方面同样有效,如Carson等人通过VR技术识别了公开报道中存在内隐群体偏见,未来可以据此进行针对性的整改以减少内隐偏见(Carson & Politte, 2021)。

词嵌入(Word Embeddings)作为自然语言处理中的关键技术,在揭示和干预内隐社会认知方面具有重要的应用价值。大规模语料训练的词向量可以反映社会中种族、性别等内隐态度和内隐刻板印象(包寒吴霜等, 2023)。例如,基于词嵌入的分析揭示了自然语言中内隐偏见的存在,研究显示白人/黑人群体在语义空间中的位置分别与“优越”或“劣等”的属性词更接近(Lee et al., 2024)。未来的研究可以利用词嵌入技术识别在线平台中的偏见性倾向和表达,并通过适当的反馈机制对用户进行引导,从而对内隐偏见进行干预。例如, Kroon通过词嵌入识别出新闻媒体报道中的内隐偏见,强调媒体有责任减少无意识的偏见,以避免对特定族裔群体的不公平刻板印象(Kroon & van der Meer, 2023)。未来的研究可以进一步开发并推广这些基于自然语言处理的工具,在大规模数据环境中自动识别内隐偏见,并结合大语言模型、情感计算、机器学习等多种人工智能技术,形成智能化、个性化的干预系统。例如,利用自然语言处理技术对个体的在线文本进行分析以识别文本中所蕴含的偏见、情感倾向等,结合机器学习算法来对个体的文本进行实时监测和动态分析从而识别个体在不同情境下的认知变化,及时发

现潜在的社会认知问题(章彦博等, 2025; Ji et al., 2020)。大语言模型可以根据识别到的信息生成个性化的干预内容,例如针对性的建议、说服性的信息或具体的行动方案等(Huang et al., 2024)。通过这些技术可以识别个体在不同情境中的内隐社会认知,自动生成合适的内隐训练方案,并不断优化干预策略,提供实时且动态的干预建议,以确保干预效果的持续性,从而实现人工智能技术实现对个体、群体行为的实时个性化干预。

此外,宏观的教育和文化在减少负面内隐社会认知方面也发挥着重要作用。教育可以通过针对性课程设计和教学活动,培养个体的多元文化意识,从而减少内隐偏见和刻板印象;文化层面则可以促进跨文化交流和文化多样性教育,增强个体对不同文化的理解和包容,进一步削弱负面内隐社会认知。这种融合技术、教育与文化干预的系统,不仅能够为人们提供实时便捷的干预,还能在教育、医疗、企业培训等实际场景中广泛应用,从而在个体和社会层面实现对内隐社会认知的有效干预。

参考文献

(*标记文献为纳入元分析的文献)

- 包寒吴霜,王梓西,程曦,苏展,杨盈,张光耀, ... 蔡华俭. (2023). 基于词嵌入技术的心理学研究:方法及应用. *心理科学进展*, 31(6), 887-904.
- *曹初. (2015). 想象性群际接触对男同性恋内外偏见的影响 [硕士学位论文]. 成都:四川师范大学.
- *陈景红. (2016). 大学生对残疾人的内隐态度研究 [硕士学位论文]. 郑州:郑州大学.
- 高可翔,张岳瑶,李思瑾,袁加锦,李红,张丹丹. (2023). 腹内侧前额叶在内隐认知重评中的因果作用. *心理学报*, 55(2), 210-223.
- *李艳敏. (2012). 想象性群际接触、接触经验对农民工内隐偏见的影响 [硕士学位论文]. 苏州:苏州大学.
- *刘璐,赵玉芳,黄雨晴,王一竹,朱勇. (2012). 提升内隐自尊减少社会排斥诱发的攻击性. *西南师范大学学报:自然科学版*, (8), 135-138.
- *刘予玲. (2010). 同性恋大学生对同性恋的外显态度和内隐态度及其干预 [硕士学位论文]. 郑州:郑州大学.
- *解诗薇. (2019). 想象性群际接触对肥胖女性内隐偏见的的影响 [硕士学位论文]. 西安:陕西师范大学.
- *施天会. (2007). 评价性条件作用改变内隐自尊的实验研究 [硕士学位论文]. 昆明:云南师范大学.
- 杨紫嫣,刘云芝,余震坤,蔡华俭. (2015). 内隐联系测验的应用:国内外研究现状. *心理科学进展*, 23(11), 1966-1980.
- 章彦博,黄峰,莫柳铃,刘晓倩,朱廷劭. (2025). 基于大

- 语言模型的自杀意念文本数据增强与识别技术. *心理学报*, 57(6), 987-1000.
- *朱睿达, 陈彩琦. (2013). 评价性条件反射对健康饮食的内隐态度、外显态度和行为的影响. *人类工效学*, 19(2), 66-70.
- *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.
- 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, 75-87.
- 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.
- Cai, H., Luo, Y. L., Chen, J., Li, X., & Xie, Y. (2021). Is implicit social cognition developmentally stable? A longitudinal study. *Developmental Psychology*, 57(12), 2220-2233.
- Cai, H., & Wu, L. (2021). The self-esteem Implicit Association Test is valid: Evidence from brain activity. *PsyCh Journal*, 10(3), 465-477.
- 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), 796-818.
- Carbone, J., & Diekmann, S. (2024). An update on recent advances in targeted memory reactivation during sleep. *NPJ Science of Learning*, 9(1), 31.
- 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.
- *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.
- *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.
- 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, 202-208.
- 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. (2021). Patterns of implicit and explicit attitudes II. Long-term change and stability, regardless of group membership. *American Psychologist*, 76(6), 851-869.
- 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), 1347-1371.
- 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.
- *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), 205–215.
- *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.
- Cooper, L. A., Saha, S., & Van Ryn, M. (2022). 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>.
- Cox, W. T., & Devine, P. G. (2019). The prejudice habit-breaking intervention: An empowerment-based confrontation approach. In R. K. Mallett & M. J. Monteith (Eds.), *Confronting prejudice and discrimination* (pp. 249–274). Academic Press.
- *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.
- *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), 359–379.
- 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.
- *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., & 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., 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., 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 STEM departments. *Journal of Experimental Social Psychology*, 73, 211–215.
- 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.
- *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.
- 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 D. D. Schmorow & C. M. Fidopiastis(Eds.), *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>.
- *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.
- 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/10.1177/10888683241-273350>.

- 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), 177–198.
- Farmer, H., Hewstone, M., Spiegler, O., Morse, H., Saifullah, A., Pan, X., ... Terbeck, S. (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), 29.
- 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.
- *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), 425–440.
- 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), 139–155.
- *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), 947–956.
- Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38, 69–119.
- *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), 123–130.
- *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), 102–113.
- 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., Fries, M., ... Wiers, R. W. (2022). Best research practices for using the Implicit Association Test. *Behavior Research Methods*, 54(3), 1161–1180.
- 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), 7–40.
- Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, 71(1), 419–445.
- 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.
- *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.
- 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), 743–760.
- *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, 50–64.
- *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), 847–851.
- *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(1), 79–86.
- 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.
- 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*, 1–8. <https://doi.org/10.1109/TCSS.2024.3475413>.
- *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>.
- *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), 350–357.
- *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., 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., 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., & 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.
- *Hunsinger, M., Christopher, M., & Schmidt, A. M. (2019). Mindfulness training, implicit bias, and force response decision-making. *Mindfulness*, 10(12), 2555–2566.
- 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.
- *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>
- *Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology*, 41(3), 137–146.
- *Kang, Y., & Falk, E. B. (2020). Neural mechanisms of attitude change toward stigmatized individuals: Temporoparietal junction activity predicts bias reduction. *Mindfulness*, 11(6), 1378–1389.
- *Kang, Y., Gray, J. R., & Dovidio, J. F. (2014). The nondiscriminating heart: Lovingkindness 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.
- 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), 424–434.
- *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), 670–675.
- 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.
- *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/s40359-023-01344-7>.
- Lieberman, M. D., Hariiri, 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), 315–326.
- 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/10.13140/2.1.1888.7681>
- 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), 797–820.
- Moran, T., Nudler, Y., & Bar-Anan, Y. (2023). Evaluative conditioning: Past, present, and future. *Annual Review of Psychology*, 74(1), 245–269.
- Narayan, M. C. (2019). CE: Addressing implicit bias in nursing: A review. *The American Journal of Nursing*, 119(7), 36–43.
- 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., ... Farné, 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), 1318–1330.

- *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.
- *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), 968–973.
- 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), 89–96.
- *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), 37–53.
- *Romano, L. E. (2025). Mindful discipline: A pilot study exploring mindfulness and the reduction of punitive discipline for Black male students. *School Psychology*, 40(3), 410–417. <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), 285–300.
- *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, 60–70.
- 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), 162–194.
- 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), 21–41.
- 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.
- *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.
- *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), 1332–1345.
- 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), 1062–1069.
- 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), 94–103.
- *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.
- *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.

- Sullivan-Bissett, E. (2023). Virtually imagining our biases. *Philosophical Psychology*, 36(4), 860–893.
- *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), 1773–1798.
- 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), 299–303.
- *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/fpsyg.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, 1–9.
- Vitriol, J. A., & Banaji, M. R. (2024). Eliciting cognitive consistency increases acceptance of implicit bias. *Journal of Experimental Psychology: General*, 153(11), 2859–2878.
- Vuletic, H. A., & Payne, B. K. (2019). Stability and change in implicit bias. *Psychological Science*, 30(6), 854–862.
- 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). Association for Computing Machinery.
- *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), 282–292.
- *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), 355–359.
- 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), 309–320.
- *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, 1–12.
- *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.
- Zhu, X., & Liao, Q. (2020). Research on implicit occupational gender stereotype based on IAT. *Advances in Psychology*, 10(2), 124–128.

Interventions for implicit social cognition

GUO Xiaoli, CHANG Junyao, SHA Maajie, YANG Ziyan

(State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology,

Chinese Academy of Sciences, Beijing 100101, China)

(Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Implicit social cognition has attracted a large number of studies. Early research primarily focused on the measurement and impact of implicit social cognition. Extensive studies have demonstrated that negative implicit social cognition, such as implicit bias, implicit stereotypes, and low implicit self-esteem, have various negative effects on individuals and society. In recent years, research has focused on in the interventions for implicit social cognition. Research has suggested nine effective intervention methods, including evaluative conditioning, approach-avoidance training, intergroup contact, exposure to counter-stereotypical exemplars, unconscious bias training, implementation intentions, targeted memory reactivation, emotion inducing, and mindfulness meditation. The current paper systematically reviews these intervention methods. The results indicate that these interventions have small to medium effect sizes, and that most intervention are effective in the short term but not in long term. Future research could combine multiple methods to enhance the long term effects. In addition, future interventions on implicit social cognition can be more efficient and convenient with the application of artificial intelligence technology.

Keywords: implicit social cognition, implicit association test, implicit bias, implicit stereotypes, psychological intervention