Chapter for *The Brains that Pull the Triggers: Syndrome E*
Editors: Itzhak Fried, Alain Berthoz, & Gretty Mirdal

Changing hearts and minds to grow compassion: Syndrome C

Heather D. Lucas[1], Daniel Sanchez[2], Jessica D. Creery[3], Xiaoqing Hu[4], & Ken A. Paller[3]

[1]Department of Psychology, Louisiana State University, Baton Rouge, Louisiana, USA
[2]Information & Computing Sciences, SRI International, Menlo Park, California, USA
[3]Department of Psychology and Cognitive Neuroscience Program, Northwestern University, Evanston, Illinois, USA
[4]Department of Psychology and State Key Laboratory of Brain and Cognitive Science, The University of Hong Kong, Pokfulam, Hong Kong

Acknowledgements
We are grateful for support from the United States National Science Foundation (BCS-1461088, BCS-1461088, BCS-1829414), the Mind and Life Institute (Varela Grant), the Mind Science Foundation, and NIH Training Grant T32-NS047987. Portions of this chapter were adapted from our recently published article on "Grappling with implicit social bias: A perspective from memory research" (Lucas et al., 2019)[1].

This book documents the challenge of understanding the human capability for extreme violence directed towards other people, as in what Itzhak Fried[2] termed *Syndrome E*. In this chapter, we wish to bring into view the opposite human capability, that of extreme compassion. Humans are capable of vicious cruelty, but also of tremendous altruism[3]. Considering both extremes can provide insights into how people can learn to avoid violent inclinations, develop habits of nonviolence, and cultivate a high regard for the welfare of others.

A hallmark of extreme compassion is arguably equanimity, or the ability to apply the same level of compassion to all, regardless of physical, psychological, or social distance. Some people so strongly exhibit this quality that they might be said to have achieved a state of compassion worthy of being dubbed *Syndrome C*—not a disorder but an ideal. The writer Marc Ian Barasch[4] documented many examples of such overwhelming compassion. Consider, for example, the extraordinary individuals who do not hesitate to donate one of their kidneys to a complete stranger. How do people manage to generate this level of compassion?

Several chapters in this book delve into the possible origins of *Syndrome E,* which can reference a variety of factors (neural, developmental, socio-political, etc.). While the origins of *Syndrome C* are likely to be equally complex and varied, there is little doubt about one thing: People can learn to be more compassionate. Accordingly, with an emphasis on learning, our goal in this chapter is to delve into forces that can move an individual away from tribalism and toward equanimity.

Our orientation relies on a fundamental neurocognitive distinction between: *(i) explicit* forces that arise due to conscious thoughts and attitudes in our minds, and *(ii) implicit* forces that operate without our awareness. Both varieties influence what we do, say, think, and feel. Our primary focus will be on implicit forces, the "dark matter" of the mind[5]. Implicit aspects of cognition may be hidden from view, but they need not remain in the dark. Understanding implicit cognition can reveal ways to work on lessening detrimental outcomes. Long-standing habits are not easily changed, particularly in the face of cultural, cognitive, and structural barriers to egalitarianism. Nonetheless, achieving a better understanding of the relevant neurocognitive mechanisms could ultimately help light the way forward.

Implicit Social Bias and Tribalism

Implicit social cognition refers to a collection of nonconscious forces that shape social behaviors, including attitudes, stereotypes, and self-evaluations[6]. Implicit social biases can be demonstrated at the individual level, such as when one unintentionally allows the gender of an individual to influence assessments of competence[7]. At the group level, social biases of implicit and explicit varieties can reinforce and magnify one another. Modern considerations such as technology-enabled mass communication may contribute additional levels of magnification by providing fast- and far-reaching avenues for the spread of prejudicial ideas[8].

Although the pervasive human tendency to distinguish ingroup from outgroup members can seem innocuous, under some circumstances such distinctions can trigger harmful responses ranging from subtle tribalism to dehumanizing instincts, potentially leading us down the path to *Syndrome E*. Individuals have an increased propensity for aggression when conflict occurs at the group level[9], and brain-imaging evidence suggests that reward-related neural responses to the successes of ingroup members are similar to those evoked by failure on the part of a rival outgroup member[10].

Importantly, the fact that implicit social biases—including those triggered by ingroup/outgroup categorization—operate without direct intentional control does not mean that we are unable to regulate these influences. Research on implicit social cognition can be considered a branch of an older body of work that distinguishes between implicit and explicit influences on human learning and memory. As such, our position is that research on domain-general principles of human memory systems provides a useful framework through which to further our understanding of the causes, consequences, and potential solutions to implicit social bias. In this spirit, we offer a review of recent and emerging points of contact. We begin by discussing parallels between these literatures that provide insight into hidden influences on our social minds. We then discuss two lines of research that focus on ways to harness plasticity with our memory systems to counter unwanted influences, perhaps moving the needle in the direction of *Syndrome C*.

The first line of research connects with the cognitive strategy of *individuation*, which involves directing attention and processing resources beyond social category memberships to focus on the characteristic features of an individual person. The second line of research concerns the potential utility of *counter-stereotype training* through repeated exposure to members of stereotyped groups in association with counter-stereotypical concepts. Previous attempts to counter implicit social biases using this type of training have found success, but only in the very short-term. Recent developments in the study of memory consolidation, and new insights about the role of sleep in consolidation, have spawned novel implications with respect to enhancing training efforts.

Multiple Memory Systems and Implicit Social Bias

Implicit memory refers to a diverse collection of processes—including cognitive and motor skill learning, habit learning, conditioning, and priming—by which traces of past experience can influence behavior without awareness of this influence[11]. The neural systems linked to implicit memory are separate from the system responsible for *declarative memory*[12,13]. Declarative memory, such as the recall and recognition of facts and episodes, is typically characterized by awareness of retrieval. Although these memory systems are separable, in practice our thoughts and behaviors—social and otherwise—are usually determined by interactions among multiple systems.

Implicit memory can be conceptualized as a form of pervasive plasticity by which virtually all of our neural systems continuously adjust to reflect statistical regularities or co-occurrences in our environments[14]. In other words, implicit memory improves processing efficiency by re-shaping the brain to reflect the environment in which processing takes place. The obligatory and continuous nature of this re-shaping includes socially constructed co-occurrences, such as stereotypic associations of social groups with certain traits, attitudes, or concepts. These stereotypes are learned early in life[15] and are reinforced through an insidious entanglement of factors that are external to members of the stereotyped groups. To point to just a few such factors, stereotypic associations between Black individuals and crime can be reinforced by biased media coverage[16], disparities in policing[17], and the assignment of longer and harsher sentences to Black versus non-Black defendants convicted of similar crimes[18].

Priming is one manifestation of implicit memory that can contribute to the acquisition and expression of social biases (see also[5]). Priming refers to altered processing due to prior experience, regardless of whether or not that experience is consciously remembered. In typical priming experiments that demonstrate preserved implicit memory in amnesic patients, a specific prior experience is shown to increase the fluency of processing for subsequently encountered information with overlapping features. Often, this fluency leads to processing that is faster and more accurate for primed relative to unprimed information. For example, the amnesic patients studied by Paller and colleagues[19] were better able to identify briefly flashed words when those words had been shown previously in the experiment, even though their conscious memory for having seen the words was impaired (**Figure 1A**). These participants also erroneously judged the primed words as having been shown longer than the unprimed words, indicating that fluency influenced perception of duration. Importantly, healthy subjects had similar priming effects, demonstrating that we are all susceptible to influences from priming.

The types of processing fluency that give rise to priming are multifaceted and can include both perceptual fluency (facilitated processing of percepts such as word forms or visual features), and fluency with meaning-based or conceptual information. Indeed, social biases as expressed on implicit association tests (IATs, which are frequently used to quantify implicit bias)[20] can be considered a form of conceptual priming, in which learned stereotypes or attitudes regarding group members are activated and facilitate processing of congruent concepts. In a typical IAT, the participant is asked to rapidly categorize stimuli that are presented centrally on a computer screen. There are usually two sets of categories, which can be denoted by anchors at the top-left and top-right of the screen, such as a social category (e.g., White-Black; Male-Female) and a stereotype-relevant category (Good-Bad; Science-Art). The key variable of interest is the extent to which categorization speed is facilitated when category anchors are versus are not aligned in a stereotype-congruent manner. With a stereotype-incongruent alignment, for example, participants would use the same motor response (e.g., pressing a key on the right side) to categorize faces as "Female" and also to categorize words as related to "Science." Because faster IAT responses are thought to reflect automated processing leading to easier categorization decisions, enhanced speed for the congruent condition relative to the incongruent condition is interpreted as resulting from an implicit association between the category and the stereotype. **Figure 1B** shows a novel example of an IAT that we developed to assess implicit bias based on political orientation.

Besides the extensive body of research with the IAT, the literature also contains numerous reports of bi-directional priming between group members and stereotype-congruent content, sometimes with serious downstream consequences. For example, exposure to Black faces reduced the perceptual threshold of clarity needed to recognize crime-relevant objects embedded in noise, suggesting that crime-related information was processed more fluently due to stereotypic links between Black individuals and crime[21]. This phenomenon has also been linked to more frequently mistaking harmless objects for weapons in the presence of Black individuals[22], which has been implicated in the unnecessary escalation of conflicts between police officers and Black individuals. These forms of priming contribute to the self-perpetuating nature of social biases by allowing them to continue and even strengthen in the absence of external reinforcement.

The consequences of social categorization and stereotyping also extend to the motor system, as evidenced by facilitated approach- and avoidance-like movements in response to images of ingroup and outgroup members, respectively[23]. Accordingly, an intriguing recent line of research has focused on counter-training this tendency by repeatedly pairing simple "approach" motor behaviors with images of racial outgroup members. In work by Kawakami and colleagues[24], White participants showed reduced anti-Black bias on an IAT after a training session that involved pulling a joystick toward oneself (indicating approach) in response to images of Black faces, and away from oneself (indicating avoidance) in response to images of White faces. Strikingly, positive effects of this training were apparent even in certain non-verbal behaviors (decreased proximal distance and more partner-focused body orientation) displayed when participants interacted with a Black confederate. This pattern implicates stimulus-response contingencies within sensorimotor systems as another potential point of entry into social bias. More generally, findings such as these illustrate that the utility of juxtaposing research on implicit social bias and memory can go beyond explanations for social phenomena to reveal potential solutions.

Limiting Implicit Bias Through Individuation

The cognitive strategy of *individuation* involves directing attention and processing resources beyond social category memberships to focus on the characteristic features of an individual. Deliberately focusing on the unique qualities of an individual not only limits the application of stereotypes, but can also reduce tendencies to dehumanize outgroup members; for example, by characterizing them as holding fewer uniquely human traits and emotions relative to ingroup members[25]. Failures to individuate are thus intimately connected to tribalism and the ensuing tendencies toward violence and aggression. Here we consider individuation specifically in the context of face perception and memory.

The general finding that *other-race* (OR) faces are not remembered as well as same-race (SR) faces is known as the *other-race effect* (ORE). The ORE can be considered a behavioral manifestation of the tendency to engage in less individuation of other-race faces. Causes of the ORE have been debated[26]. *Perceptual expertise* accounts suggest that a lack of interracial contact results in perceptual systems that are poorly tuned to the physical dimensions along which OR faces tend to differ from one another[27,28]. In contrast, *social-cognitive accounts* point to factors such as a lower motivation to individuate OR faces and a tendency to instead focus on race-specifying information[29,30]. Many contemporary accounts of the ORE are hybrid models that posit *perceptual-social linkages*, or interactions between perceptual and social elements of face processing[31,34]. Such models count perceptual expertise as one of many factors that impact the extent to which a face is initially encoded in relation to ways in which it is distinct from other faces in the same social category, which, in turn, impacts the activation of category-relevant stereotypes and attitudes[35].

An important implication of these models is that the ORE should not be viewed as a perceptual phenomenon that is epiphenomenal to social biases. Instead, failures of individuation that produce the ORE constitute yet another way in which biases are self-perpetuating. As such, research aimed at understanding the mechanisms of the ORE—and factors that help to reduce it—can be included as part of comprehensive efforts to contend with social bias.

In theory, face processing at any point during a social interaction may shift between individuation and categorization. However, event-related potential (ERP) studies have revealed that social-categorical information can permeate the very earliest stages of face processing. Two brain potentials in particular—an occipitotemporal N250 and a frontocentral N200—have been identified as candidate markers of the preferential early individuation of faces belonging to ingroup members. Understanding these neural signatures of individuation may point us toward effective intervention strategies that influence perceptual processing stages critical for OR face encoding.

The N250 is a bilateral negative component that peaks ~250 ms after the onset of a visual stimulus, and is particularly pronounced for faces and other stimulus categories for which the perceiver has expertise differentiating among exemplars at the subordinate level[36]. Interestingly, one study that combined ERPs with expertise training[37] found that the effectiveness of a perceptual training procedure designed to improve OR face recognition correlated across subjects with the magnitude of increases in N250 amplitudes for OR faces. The frontocentral N200 is a midline negative component that is larger (more negative) for SR than for OR faces[38-40]. The N200 appears to be functionally dissociable from the N250, in that the former is sensitive to race-based attentional biases that cannot be directly attributed to perceptual expertise. For example, both Black and White faces yielded greater N200 amplitudes in White participants when those faces were preceded by descriptions of stereotype-incongruent versus stereotype-congruent behaviors, the former of which may have encouraged attention to individuating features[41].

Work in our laboratory[40] has directly related N200 attentional biases in White participants to the ORE by using the *subsequent-memory technique* to shed light on how encoding failures for OR faces differed qualitatively from those for SR faces (**Figure 1C**). The subsequent-memory technique involves comparing brain activity during initial encoding for stimuli that are subsequently remembered versus stimuli that are subsequently forgotten[42]. These comparisons reveal *d*ifferential neural activity based on *m*emory (*Dm effects*), which index encoding operations that are pivotal in determining whether information will be retained in long-term memory. A key finding from our work[40] was that a strikingly early subsequent memory effect occurred on N200 potentials that was specific to OR faces. Specifically, N200 potentials were more negative for OR faces that were remembered relative to those that were forgotten, suggesting that the fate of OR face memory hinged on a very early processing stage associated with individuation. Importantly, failures to individuate OR faces at this early stage may preclude the effective use of downstream memorization strategies related to conceptual individuation, thus shifting processing toward categorization and away from individuation—and, unfortunately, away from humanization.

A key implication of this work is that interventions that focus on developing expertise with OR face individuation should be taken seriously as a potential route for reducing the negative impact of social categorization and its sequelae (see[35,43,44] for examples of interventions that successfully used these techniques to reduce implicit social bias). Such efforts may be combined with other techniques that have been found to promote humanization, such as emphasizing ways in which outgroup members can be categorized on a variety of dimensions rather than just one[25].

Countering Implicit Bias by Hacking into Memory During Sleep

Another line of relevant research concerns the ways that experiences can directly induce changes in implicit social bias though counter-stereotype training. In counter-stereotype training, links between members of a group and their stereotypical characteristics are weakened. This sort of training can be effective in temporarily reducing implicit social biases when measured immediately after training, likely because the training reinforces counter-stereotypical associations and renders them more accessible[45,46]. Insights from recent research on memory—particularly concerning the processes that promote the stabilization of new memories—have spawned novel implications with respect to enhancing and prolonging the effects of such training efforts.

Although initial encoding is necessary for memory, it is not sufficient. As time passes, new knowledge is likely to be forgotten unless it is consolidated. *Consolidation* refers to the process that transforms and stabilizes newly encoded memories as they become integrated with other memories[47]. Consolidation of declarative memories at a neural-systems level is thought to involve hippocampal-neocortical interaction. Repeated reactivation of the same neural circuits that contributed to initial encoding seems to drive a gradual reorganization, allowing links to be formed and strengthened in distributed cortical networks. Indeed, memory reactivation may be compulsory if a memory is to endure beyond a short period of time[48]. Although consolidation is most often discussed in relation to declarative memory, systems-level changes can also influence learning within implicit memory systems.

Sleep is also relevant, as various research findings have implicated sleep in memory consolidation[49,50]. The observation that hippocampal place cells replay learning-related firing patterns during sleep[51] prompted many investigators to hypothesize that sleep-based reactivation of memory traces are critical to the consolidation process. Mounting evidence points to *slow-wave sleep* (SWS) as particularly important, perhaps due to the role that slow oscillations play in timelocking thalamo-cortical spindles, which in turn coordinate the timing of hippocampal sharp-wave ripples[52]. These nested oscillations conceivably provide a vehicle for the informational interactions across brain regions that are necessary for memory consolidation.

In 2007, Rasch and colleagues[53] made the key discovery that odor-cued reactivation could be used as a tool for biasing consolidation during sleep. They succeeded in strengthening spatial memories by 1) associating an odor with the learning context in which spatial memories were acquired, and 2) delivering the same odor during overnight periods of SWS. Subsequent research by our group and others suggests that this reactivation can be used in a targeted manner, effectively singling out specific memories for consolidation over others. For example, Rudoy and colleagues[54] modified Rasch's paradigm such that, rather than pairing an odor with an entire learning context for object-location associations, a unique sound was paired with *each* learned association By playing only a subset of these sounds during a post-encoding period of SWS, we were able to selectively improve subsequent memory for cued relative to uncued object-location associations.

This strategy, termed *targeted memory reactivation* (TMR), extends well beyond spatial memory[55-59]. Of particular relevance, Hu and colleagues[59] obtained evidence that manipulating memory processing during sleep can facilitate training that relies on exposure to counter-stereotypical information (**Figure 1D**). In previous work, bias reduction resulting from such procedures has tended to be disappointingly fleeting, with even strong immediate effects dissipating within 1-2 days[60]. Given that the long-term viability of these training effects likely depends on consolidation, it is possible that training can be enhanced using a manipulation to promote consolidation, such as TMR.

As depicted in **Figure 1D**, Hu and colleagues[59] conducted an initial test of this hypothesis using outcome measures based on a race IAT (Black-White) and a gender IAT (which tested preferential associations of males with science and women with art compared to the reverse associations). After collecting baseline bias scores, counter-stereotype training procedures were administered for both of these biases, with the addition of two novel sounds that were each selectively associated with one of the two training contexts. Then, in a post-training nap, one of the two sounds was quietly presented multiple times during SWS.

The chief prediction in this experiment was that the sounds presented during SWS would differentially influence the two bias scores. This prediction was born out. After the nap, bias levels differed; they were further reduced *only* for the bias that was targeted during SWS. These results are consistent with the notion that, during sleep, sound-induced reactivation of the learning context in which training occurred served to amplify the effects of that training. Perhaps most strikingly, TMR increased the longevity of counter-stereotype training; IAT scores at 1 week compared to immediately after training were disproportionately changed in accordance with which sound was presented during sleep.

Some limitations of this experiment warrant mention. First, the extent to which implicit measures such as the IAT can reliably predict overt discrimination is a subject of ongoing debate[61-65]. As such, the results do not tell us whether behavior in actual interpersonal interactions would be less racist or sexist as a result of TMR. Second, although we were able to find training effects even after 1 week, with significant effects of TMR on pre-post IAT differences, these effects were quite small. It is possible that effect sizes could scale with additional TMR or training sessions, given that Hu et al. employed only a single, 10-minute training session. However, this speculation awaits empirical evidence.

Additional research is also needed to evaluate the usefulness of such procedures outside of the laboratory. The extent to which an individual's ordinary actions are influenced by implicit social bias is very difficult to measure with precision. On the other hand, strategies can be applied to limit such influences[64], and one can also take steps to reduce implicit social bias through learning. The memory research we have featured here underscores the notion that bias formation and bias reduction adhere to domain-general principles of memory processing, which manifest both during wake and sleep. The findings of Hu and colleagues[59] call attention in particular to the role of consolidation in bias modification. Other principles derived from memory research, including the well-established benefits of engaging in spaced rather than massed practice, may also be relevant for efforts to reduce unwanted biases through learning. Importantly, while such efforts may one day be imperative for people in positions of power (such as judges, police officers, and those making hiring decisions), the reliance on basic memory processing should render these strategies available to anyone who is willing to devote time and effort toward equanimity.

A Way Forward, Both Implicitly and Explicitly

Altering a social bias is a far more complex undertaking than, for example, learning a new motor skill or correcting a factual misconception. Social influences are ubiquitous. We are bombarded daily by information that promotes the biases perpetuated in contemporary society. Some even suggest that we are evolutionarily pre-disposed to grasp for ingroup/outgroup distinctions[67].

On the other hand, research on learning and memory continues to underscore the human capacity for change. Human brains are remarkably plastic, and we can envision a future version of our society that does better by taking a sustained and proactive approach to harnessing this plasticity for prosocial goals. One important component of equanimity training is the reduction of biased thoughts and behaviors. This goal can be advanced through intentional exposure to information that challenges learned stereotypes and prejudices, as well as by practicing humanizing strategies such as individuation. The advent of advocacy for proper handwashing and general cleanliness could only work once our understanding of germ theory had progressed. In the same way, neurocognitive theories can serve as starting points for practices that promote good care and maintenance of our social minds. Knowledge about the operation of our memory systems can provide insight into discrepancies between our values and our implicit knowledge, and can uncover new strategies for reducing these discrepancies.

As we continue to develop our understanding of implicit bias, we should keep in mind that the path forward may be challenging and unpredictable. Whereas this area of research has been successfully used to guide policy[68], attempts to counter biases with seemingly intuitive strategies do not always go as planned. For example, attempting to suppress stereotype-based thoughts can lead to rebound effects, in which the stereotypic information becomes more, rather than less, accessible[69–71]. Similarly, providing counter examples to compete with biased viewpoints can push people even further towards polarization[72]. Effort and ingenuity are needed to change old habits.

Current theories of human memory posit that multiple learning mechanisms come together to produce, sustain, and activate such habits. Effectively countering unwanted effects of bias on behavior will also require a multiple-systems orientation. We can take a proactive approach to the information we consume and the processing that we apply to outgroup members, bearing in mind both our fast, lazy system that gobbles up information outside of reason[73] and our slow, deliberative system that allows for a mindful interpretation[74].

Many societal and biological factors, as summarized in the present volume, have the unfortunate effect of encouraging people to trod down the path to negativity and tribalism—occasionally culminating in the extreme as *Syndrome E*. Instead, people can choose to learn and cultivate kindness. Making progress on the path leading to high levels of compassion—even seeking to attain *Syndrome C*—is not only possible for individuals, but it is also essential for society as a whole.

References

1. Lucas HD, Creery JD, Hu X, Paller KA, « Grappling with implicit social bias: A perspective from memory research », Neuroscience, 2019, 406, 684–697. doi: 10.1016/j.neuroscience.2019.01.037..
2. Fried I, « Syndrome E », The Lancet, 1997, 350, p. 1845–1847.
3. Ricard M, Altruism: The Power of Compassion to Change Yourself and the World. UK: Hachette.
4. Barasch M, Field Notes on the Compassionate Life: A Search for the Soul of Kindness. Emmaus, PA, US: Rodale Press.
5. Amodio DM, Ratner KG, « A memory systems model of implicit social cognition », Current Directions in Psychological Science, 2011, 20, p. 143–148.
6. Greenwald AG, Banaji MR, « Implicit social cognition: attitudes, self-esteem, and stereotypes. », Psychological review, 1995, 102, p. 4–27.
7. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J, « Science faculty's subtle gender biases favor male students », Proceedings of the National Academy of Sciences, 2012, 109, p. 16474–16479.
8. Hsueh M, Yogeeswaran K, Malinen S, « 'Leave your comment below': Can biased online comments influence our own prejudicial attitudes and behaviors? », Human Communication Research, 2015, 41, p. 557–576.
9. Meier BP, Hinsz VB, « A comparison of human aggression committed by groups and individuals: An interindividual–intergroup discontinuity », Journal of Experimental Social Psychology, 2004, 40, p. 551–559.
10. Cikara M, Botvinick MM, Fiske ST, « Us versus Them: Social identity shapes neural responses to intergroup competition and harm », Psychological science, 2011, 22.
11. Schacter DL, « Implicit memory: History and current status. », Journal of Experimental Psychology: Learning, Memory, and Cognition, 1987, 13, p. 501–518.
12. Eichenbaum H, Cohen NJ, From Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford psychology series; no. 35.
13. Squire LR, « Memory systems of the brain: A brief history and current perspective », Neurobiology of Learning and Memory, 2004, 82, p. 171–177.
14. Reber PJ, « The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research », Neuropsychologia, 2013, 51, p. 2026–2042.
15. Dunham Y, Chen EE, Banaji MR, « Two signatures of implicit intergroup attitudes », Psychological Science, 2013, 24, p. 860–868.
16. Dixon TL, Daniel L, « Race and the misrepresentation of victimization on local television news », Communication Research, 2000, 27, p. 547–573.
17. Epp CR, Maynard-Moody S, Haider-Markel D, « Beyond profiling: The institutional sources of racial disparities in policing », Public Administration Review, 2017, 77, p. 168–178.

18. Steffensmeier D, Ulmer J, Kramer J, « The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male », Criminology, 1998, 36, p. 763–798.

19. Paller KA, Mayes AR, McDermott M, Pickering AD, Meudell PR, « Indirect measures of memory in a duration-judgement task are normal in amnesic patients. », Neuropsychologia, 1991, 29, p. 1007–1018.

20. Greenwald AG, McGhee DE, Schwartz JLK, « Measuring individual differences in implicit cognition: The implicit association test. », Journal of Personality and Social Psychology, 1998, 74, p. 1464–1480.

21. Eberhardt JL, Goff PA, Purdie VJ, Davies PG, « Seeing Black: Race, crime, and visual processing. », Journal of Personality and Social Psychology, 2004, 87, p. 876–893.

22. Payne BK, « Weapon Bias », Current Directions in Psychological Science, 2006, 15, p. 287–291.

23. Paladino M-P, Castelli L, « On the immediate consequences of intergroup categorization: Activation of approach and avoidance motor behavior toward ingroup and outgroup members », Personality and Social Psychology Bulletin, 2008, 34, p. 755–768.

24. Kawakami K, Phills CE, Steele JR, Dovidio JF, « (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. », Journal of Personality and Social Psychology, 2007, 92, p. 957–971.

25. Prati F, Crisp RJ, Meleady R, Rubini M, « Humanizing outgroups through multiple categorization », Personality and Social Psychology Bulletin, 2016, 42, p. 526–539.

26. Meissner CA, Brigham JC, « Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review », Psychology, Public Policy, and Law, 2001, 7, p. 3–35.

27. Chiroro P, Valentine T, « An investigation of the contact hypothesis of the own-race bias in face recognition », The Quarterly Journal of Experimental Psychology Section A, 1995, 48, p. 879–894.

28. Walker PM, Hewstone M, « A perceptual discrimination investigation of the own-race effect and intergroup experience », Applied Cognitive Psychology, 2006, 20, p. 461–475.

29. MacLin OH, Malpass RS, « The ambiguous-race face illusion », Perception, 2003, 32, p. 249–252.

30. Hugenberg K, Sacco DF, « Social categorization and stereotyping: How social categorization biases person perception and face memory », Social and Personality Psychology Compass, 2008, 2, p. 1052–1072.

31. Young SG, Hugenberg K, « Individuation motivation and face experience can operate jointly to produce the own-race bias », Social Psychological and Personality Science, 2012, 3, p. 80–87.

32. Wan L, Crookes K, Reynolds KJ, Irons JL, McKone E, « A cultural setting where the other-race effect on face recognition has no social–motivational

component and derives entirely from lifetime perceptual experience », Cognition, 2015, 144, p. 91–115.

33. Hugenberg K, Young SG, Bernstein MJ, Sacco DF, « The categorization-individuation model: An integrative account of the other-race recognition deficit. », Psychological Review, 2010, 117, p. 1168–1187.
34. Anzures G, Quinn PC, Pascalis O, Slater AM, Lee K, « Development of own-race biases », Visual Cognition, 2013, 21, p. 1165–1182.
35. Lebrecht S, Pierce LJ, Tarr MJ, Tanaka JW, Ochsner K, « Perceptual other-race training reduces implicit racial bias », PLoS ONE, 2009, 4, p. e4215.
36. Scott LS, Tanaka JW, Sheinberg DL, Curran T, « A reevaluation of the electrophysiological correlates of expert object processing », Journal of Cognitive Neuroscience, 2006, 18, p. 1453–1465.
37. Tanaka JW, Pierce LJ, « The neural plasticity of other-race face recognition », Cognitive, Affective, & Behavioral Neuroscience, 2009, 9, p. 122–131.
38. Kubota JT, Ito TA, « Multiple cues in social perception: The time course of processing race and facial expression. », Journal of Experimental Social Psychology, 2007, 43, p. 738–752.
39. Dickter CL, Bartholow BD, « Racial ingroup and outgroup attention biases revealed by event-related brain potentials. », Social Cognitive and Affective Neuroscience, 2007, 2, p. 189–198.
40. Lucas HD, Chiao JY, Paller KA, « Why some faces won't be remembered: Brain potentials illuminate successful versus unsuccessful encoding for same-race and other-race faces », Frontiers in Human Neuroscience, 2011, 5, p. 20.
41. Dickter CL, Gyurovski I, « The effects of expectancy violations on early attention to race in an impression-formation paradigm », Social Neuroscience, 2012, 7, p. 240–251.
42. Paller KA, Wagner AD, « Observing the transformation of experience into memory. », Trends in Cognitive Sciences, 2002, 6, p. 93–102.
43. Xiao WS, Fu G, Quinn PC, Qin J, Tanaka JW, Pascalis O, et al., « Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children », Developmental Science, 2015, 18, p. 655–663.
44. Qian MK, Quinn PC, Heyman GD, Pascalis O, Fu G, Lee K, « Perceptual individuation training (but not mere exposure) reduces implicit racial bias in preschool children. », Developmental Psychology, 2017, 53, p. 845–859.
45. Olson MA, Fazio RH, « Reducing automatically activated racial prejudice through implicit evaluative conditioning », Personality and Social Psychology Bulletin, 2006, 32, p. 421–433.
46. Gawronski B, LeBel EP, « Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not », Journal of Experimental Social Psychology, 2008, 44, p. 1355–1361.
47. Squire LR, Genzel L, Wixted JT, Morris RG, « Memory consolidation », Cold Spring Harbor Perspectives in Biology, 2015, 7, p. a021766.

48. Paller KA, Antony JW, Mayes A, Norman KA, « Replay-based consolidation governs enduring memory storage », In: Gazzaniga MS, Mangun GR, Poeppel D, editors. The Cognitive Neurosciences, 6th ed. MIT Press.

49. Oudiette D, Paller KA, « Upgrading the sleeping brain with targeted memory reactivation. », Trends in Cognitive Sciences, 2013, 17, p. 142–149.

50. Rasch B, Born J, « About sleep's role in memory. », Physiological Reviews, 2013, 93, p. 681–766.

51. Wilson MA, McNaughton BL, « Reactivation of hippocampal ensemble memories during sleep. », Science, 1994, 265, p. 676–679.

52. Staresina BP, Bergmann TO, Bonnefond M, van der Meij R, Jensen O, Deuker L, et al., « Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. », Nature Neuroscience, 2015, 18, p. 1679–1686.

53. Rasch B, Buchel C, Gais S, Born J, « Odor cues during slow-wave sleep prompt declarative memory consolidation », Science, 2007, 315, p. 1426–1429.

54. Rudoy JD, Voss JL, Westerberg CE, Paller KA, « Strengthening individual memories by reactivating them during sleep », Science, 2009, 326, p. 1079.

55. Batterink LJ, Oudiette D, Reber PJ, Paller KA, « Sleep facilitates learning a new linguistic rule », Neuropsychologia, 2014, 65, p. 169–179.

56. Batterink LJ, Paller KA, « Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation », Brain and Language, 2017, 167, p. 83–93.

57. Honma M, Plass J, Brang D, Florczak SM, Grabowecky M, Paller KA, « Sleeping on the rubber-hand illusion: Memory reactivation during sleep facilitates multisensory recalibration », Neuroscience of Consciousness, 2016, 2016, p. niw020.

58. Antony JW, Gobel EW, O'Hare JK, Reber PJ, Paller KA, « Cued memory reactivation during sleep influences skill learning », Nature Neuroscience, 2012, 15, p. 1114–1116.

59. Hu X, Antony JW, Creery JD, Vargas IM, Bodenhausen G V, Paller KA, « Unlearning implicit social biases during sleep », Science, 2015, 348.

60. Lai CK, Skinner AL, Cooley E, Murrar S, Brauer M, Devos T, et al., « Reducing implicit racial preferences: II. Intervention effectiveness across time », Journal of Experimental Psychology: General, 2016, 145, p. 1001–1016.

61. Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR, « Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity », Journal of Personality and Social Psychology, 2009, 97, p. 17–41.

62. Oswald FL, Mitchell G, Blanton H, Jaccard J, Tetlock PE, « Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies », Journal of Personality and Social Psychology, 2013, 105, p. 171–192.

63. Greenwald AG, Banaji MR, Nosek BA, « Statistically small effects of the Implicit Association Test can have societally large effects », Journal of Personality and Social Psychology, 2015, 108, p. 553–561.

64. Payne BK, Vuletich HA, Lundberg KB, « The bias of crowds: How implicit bias bridges personal and systemic prejudice », Psychological Inquiry, 2017, 28, p. 233–248.

65. Gawronski B, Bodenhausen G V, « Beyond persons and situations: An interactionist approach to understanding implicit bias », Psychological Inquiry, 2017, 28, p. 268–272.

66. Singletary SL, Hebl MR, « Compensatory strategies for reducing interpersonal discrimination: The effectiveness of acknowledgments, increased positivity, and individuating information », Journal of Applied Psychology, 2009, 94, p. 797–805.

67. Kurzban R, Tooby J, Cosmides L, « Can race be erased? Coalitional computation and social categorization », Proceedings of the National Academy of Sciences, 2001, 98, p. 15387–15392.

68. Payne BK, Vuletich HA, « Policy insights from advances in implicit bias research », Policy Insights from the Behavioral and Brain Sciences, 2018, 5, p. 49–56.

69. Macrae CN, Bodenhausen GV, Milne AB, Wheeler V, « On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory », Social Cognition, 1996, 14, p. 1–20.

70. Macrae CN, Bodenhausen GV, Milne AB, Jetten J, « Out of mind but back in sight: Stereotypes on the rebound », Journal of Personality and Social Psychology, 1994, 67, p. 808–817.

71. Monteith MJ, Lybarger JE, Woodcock A, « Schooling the cognitive monster: The role of motivation in the regulation and control of prejudice », Social and Personality Psychology Compass, 2009, 3, p. 211–226.

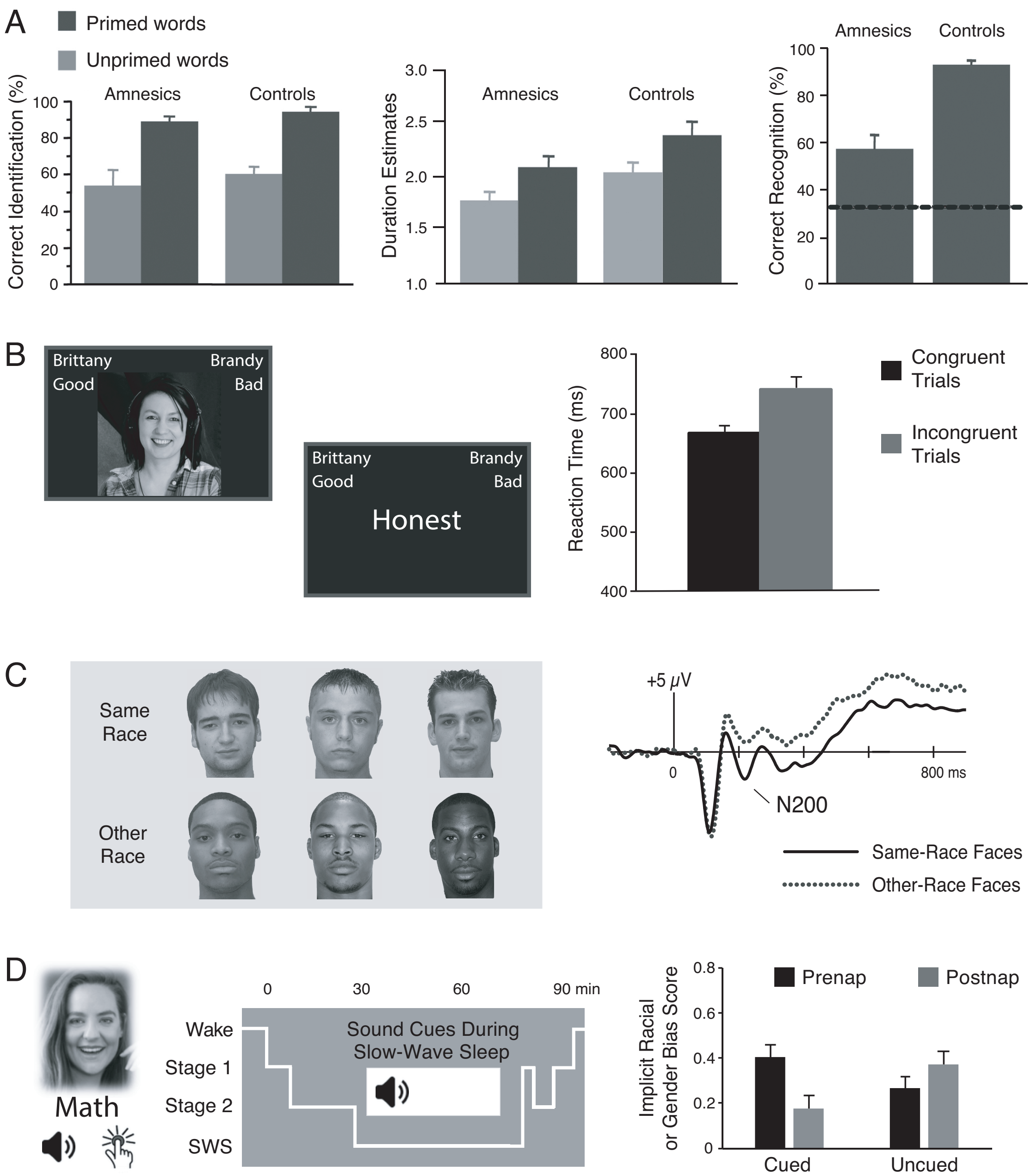
72. Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, et al., « Exposure to opposing views on social media can increase political polarization », Proceedings of the National Academy of Sciences, 2018, 115, p. 9216–9221.

73. Pennycook G, Rand DG, « Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning », Cognition, 2019, doi: 10.1016/j.neuroscience.2019.01.037. [Epub ahead of print].

74. Kahneman D, Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

75. Creery JD, Florczak SM, Zaheed AB, Antony JW, Paller KA, « An implicit measure of pro-social attitudes: Can a liberal feel equanimity toward a conservative? » Poster presented at the Cognitive Neuroscience Society annual meeting, Boston, Massachusetts, 2014.

Figure 1. Examples of ways in which *implicit* or nonconscious cognitive processes shape thoughts, attitudes, and behaviors. **(A)** The phenomenon of preserved priming in patients with amnesia indicates that implicit memory can operate independently of explicit or conscious memory. In one such study[19], results were compared between a group of 9 people with amnesia and a matched group of 9 healthy participants. In the study phase, participants were asked to read a set of words. A few minutes later, they viewed 40 primed words (from the study list) and 40 unprimed words (new words). Each word was preceded and followed by a 500-ms mask and the word duration was adjusted to be near the threshold for reading for each individual. *Left panel:* Word identification was superior for primed words compared to unprimed words in both groups, demonstrating preserved priming (implicit memory). *Center panel*: Both groups also showed priming of duration estimates, which were longer for primed words compared to unprimed words, presumably because reading fluency was mistakenly attributed to word duration. *Right panel:* As expected, amnesic participants were impaired in their ability to recognize words from the study phase (explicit memory). **(B)** An example of an Implicit Association Test (IAT) indicating bias against people with differing political views. *Left panel:* For each of two political issues, a view expressing the position the participant held was ascribed to one person and a sharply opposing position to another person. These two faces were then used in the IAT along with words corresponding to positive and negative personal qualities (e.g., intelligent, rude). *Right panel:* The expected bias was evident in the form of reaction-time differences as a function of whether response assignments were congruent (e.g., positive qualities aligned with same-position faces) or incongruent (positive qualities aligned with opposite-position faces). **(C)** Illustration of the experimental methods and results from Lucas et al[40]. *Left panel:* White participants were asked to commit both SR/White faces and OR/Black faces to memory in anticipation of a subsequent memory test. *Right panel*: Larger N200 potentials were elicited by SR relative to OR faces, consistent with greater individuation of SR faces. **(D)** Targeted memory reactivation used with training to reduce implicit social bias, from Hu et al[59]. *Left panel:* Procedure for counter-stereotype training. Specific sounds were paired with counter-stereotype training contexts for race and gender stereotypes. *Center panel:* Procedures for the nap phase of the experiment in which one of the two sounds was repeatedly played during slow-wave sleep. *Right panel:* The change in implicit bias from prenap to postnap diverged as a function of cueing condition, showing a further reduction only for the cued social bias.



A
Primed words
Unprimed words
Amnesics
Controls
Correct Identification (%)
100
80
60
40
20
0
3.0
2.5
2.0
1.5
1.0
Duration Estimates
Correct Recognition (%)
B
Brittany
Brandy
Good
Bad
Honest
800
700
600
500
400
Reaction Time (ms)
Congruent Trials
Incongruent Trials
C
Same Race
Other Race
+5 μV
0
800 ms
N200
Same-Race Faces
Other-Race Faces
D
Math
0
30
60
90 min
Wake
Stage 1
Stage 2
SWS
Sound Cues During Slow-Wave Sleep
Implicit Racial or Gender Bias Score
0.8
0.6
0.4
0.2
0
Prenap
Postnap
Cued
Uncued