However, direct activation of DG engram cells of the ANI group elicited as much freezing behavior as did the activation of these cells of the SAL group. This unexpected finding is supported by a set of additional cellular and behavioral experiments. Whereas amygdala engram cell reactivation upon exposure to the conditioned context is significantly lower in the ANI group as compared with the SAL group, optogenetic activation of DG engram cells results in normal reactivation of downstream CA3 and BLA engram cells (Fig. 5). At the behavioral level, the amnesia rescue was observed under a variety of different conditions in which one or more parameters were altered (Figs. 2 and 3 and figs. S9 and S13). Thus, our overall findings indicate that memory engrams survive a posttraining administration of protein synthesis inhibitors during the consolidation window and that the memory remains retrievable by means of ChR2-mediated direct engram activation even after retrograde amnesia is induced. The drive initiated with light-activation of one component of a distributed memory engram (such as that in the DG) is sufficient to reactivate engrams in downstream regions (such as that in CA3 and BLA) that would also be affected by the systemic injection of a protein synthesis inhibitor (ANI).

These findings suggest that although a rapid increase of synaptic strength is likely to be crucial during the encoding phase, the augmented synaptic strength is not a crucial component of the stored memory (35–37). This perspective is consistent with a recent study showing that an artificial memory could be reversibly disrupted by depression of synaptic strength (38). On the other hand, persistent and specific connectivity of engram cells that we find between DG engram cells and downstream CA3 or BLA engram cells in both SAL and ANI groups may represent a fundamental mechanism of memory information storage (39). Our findings also suggest that the primary role of augmented synaptic strength during and after the consolidation phase may be to provide natural recall cues with efficient access to the soma of engram cells for their reactivation and, hence, recall.

The integrative memory engram-based approach used here for parsing memory and amnesia into encoding, consolidation, and retrieval aspects may be of wider use to other experimental and clinical cases of amnesia, such as Alzheimer's disease (40).

REFERENCES AND NOTES

1. G. E. Müller, A. Pilzecker, *Z. Psychol.* **1**, 1–288 (1900).
2. C. P. Duncan, *J. Comp. Physiol. Psychol.* **42**, 32–44 (1949).
3. J. L. McGaugh, *Science* **287**, 248–251 (2000).
4. J. B. Flexner, L. B. Flexner, E. Stellar, *Science* **141**, 57–59 (1963).
5. L. B. Flexner, J. B. Flexner, R. B. Roberts, *Science* **155**, 1377–1383 (1967).
6. H. P. Davis, L. R. Squire, *Psychol. Bull.* **96**, 518–559 (1984).
7. E. R. Kandel, *Science* **294**, 1030–1038 (2001).
8. R. J. Kelleher 3rd, A. Govindarajan, S. Tonegawa, *Neuron* **44**, 59–73 (2004).
9. T. Takeuchi, A. J. Duszkiewicz, R. G. Morris, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130288 (2014).
10. A. Govindarajan, I. Israely, S. Y. Huang, S. Tonegawa, *Neuron* **69**, 132–146 (2011).
11. M. Krug, B. Lössner, T. Ott, *Brain Res. Bull.* **13**, 39–42 (1984).
12. U. Frey, M. Krug, K. G. Reymann, H. Matthies, *Brain Res.* **452**, 57–65 (1988).
13. Y. Y. Huang, P. V. Nguyen, T. Abel, E. R. Kandel, *Learn. Mem.* **3**, 74–85 (1996).
14. R. Semon, *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens* (Wilhelm Engelmann, Leipzig, 1904).
15. S. A. Josselyn, *J. Psychiatry Neurosci.* **35**, 221–228 (2010).
16. X. Liu et al., *Nature* **484**, 381–385 (2012).
17. S. Ramirez et al., *Science* **341**, 387–391 (2013).
18. R. L. Redondo et al., *Nature* **513**, 426–430 (2014).
19. C. A. Denny et al., *Neuron* **83**, 189–201 (2014).
20. K. Z. Tanaka et al., *Neuron* **84**, 347–354 (2014).
21. K. K. Cowansage et al., *Neuron* **84**, 432–441 (2014).
22. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).
23. R. L. Clem, R. L. Huganir, *Science* **330**, 1108–1112 (2010).
24. A. Suzuki et al., *J. Neurosci.* **24**, 4787–4795 (2004).
25. B. N. Armbruster, X. Li, M. H. Pausch, S. Herlitze, B. L. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5163–5168 (2007).
26. J. H. Han et al., *Science* **323**, 1492–1496 (2009).
27. J. R. Misanin, R. R. Miller, D. J. Lewis, *Science* **160**, 554–555 (1968).
28. K. Nader, G. E. Schafe, J. E. Le Doux, *Nature* **406**, 722–726 (2000).
29. X. Liu, S. Ramirez, S. Tonegawa, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130142 (2014).
30. A. Besnard, S. Laroche, J. Caboche, *Brain Struct. Funct.* **219**, 415–430 (2014).
31. J. Hall, K. L. Thomas, B. J. Everitt, *Eur. J. Neurosci.* **13**, 1453–1458 (2001).
32. J. L. McGaugh, *Science* **153**, 1351–1358 (1966).
33. Y. Dudai, *Annu. Rev. Psychol.* **55**, 51–86 (2004).
34. J. P. Johansen, C. K. Cain, L. E. Ostroff, J. E. LeDoux, *Cell* **147**, 509–524 (2011).
35. R. R. Miller, L. D. Matzel, *Learn. Mem.* **13**, 491–497 (2006).
36. C. A. Miller, J. D. Sweatt, *Learn. Mem.* **13**, 498–505 (2006).
37. S. Chen et al., *eLife* **3**, e02844 (2014).
38. S. Nabavi et al., *Nature* **511**, 348–352 (2014).
39. D. O. Hebb, *The Organization of Behavior; A Neuropsychological Theory* (Wiley, New York, 1949).
40. S. Daumas et al., *Learn. Mem.* **15**, 625–632 (2008).

COGNITIVE NEUROSCIENCE

Unlearning implicit social biases during sleep

Xiaoqing Hu,[1,2] James W. Antony,[1,3] Jessica D. Creery,[1] Iliana M. Vargas,[1] Galen V. Bodenhausen,[1] Ken A. Paller[1]*

Although people may endorse egalitarianism and tolerance, social biases can remain operative and drive harmful actions in an unconscious manner. Here, we investigated training to reduce implicit racial and gender bias. Forty participants processed counterstereotype information paired with one sound for each type of bias. Biases were reduced immediately after training. During subsequent slow-wave sleep, one sound was unobtrusively presented to each participant, repeatedly, to reactivate one type of training. Corresponding bias reductions were fortified in comparison with the social bias not externally reactivated during sleep. This advantage remained 1 week later, the magnitude of which was associated with time in slow-wave and rapid-eye-movement sleep after training. We conclude that memory reactivation during sleep enhances counterstereotype training and that maintaining a bias reduction is sleep-dependent.

Social interactions are often fraught with bias. Our preconceptions about other people can influence many types of behavior. For example, documented policing errors have repeatedly shown the potential harm of racial profiling (1). In experiments that used a first-person-shooter videogame, both White and Black participants were more likely to shoot Black than White individuals, even when they held a harmless object rather than a gun (2). When hiring potential research assistants, both male and female faculty members were more likely to hire male than equally qualified female candidates (3).

Although the tendency for people to endorse racist or sexist attitudes explicitly has decreased in recent years (4), social biases may nevertheless influence people's behavior in an implicit or unconscious manner, regardless of their intentions or efforts to avoid bias (5). Ample evidence indicates that implicit biases can drive discriminatory behaviors and exacerbate intergroup conflict (5–8). For instance, implicit racial biases decrease investments given to racial out-group members

[1]Department of Psychology, Northwestern University, Evanston, IL 60208, USA. [2]Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA. [3]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA.
*Corresponding author. E-mail: kap@northwestern.edu

in a trust game (6). At a broader level, the gender gap in science achievement in a nation is correlated with the level of implicit stereotyping of females as not having an aptitude for science (8).

Whereas discriminatory behaviors can be detrimental to individuals and society, implicit social bias can be difficult to correct because of a range of affective, cognitive, motivational, and social factors, as follows (9, 10). First, out-group members can be perceived as threatening, and the fear response to those individuals can resist extinction (11). Second, biases are acquired over many years of exposure to stereotypes, and they can efficiently operate without occupying cognitive resources (5, 10). Third, motivation to seek higher status or self-enhancement commonly results in out-group derogation (9, 10). Last, perceived social norms can prescribe people's expression of stereotyping and prejudice (12). Despite such challenges, implicit biases can be reduced by learning about counterstereotype cases (13). However, benefits of this counterbias training can be fragile, subject to reversal when the original stereotypes are again reinforced in typical circumstances, such as through the media (14). Longer-term reductions in implicit social biases may necessitate that counterbias training be followed by further memory consolidation, as is the case for many other types of learning (15).

Recent findings suggest that memory consolidation during sleep may be essential for preserving newly acquired information, such as declarative and procedural memories (15–19). During sleep, information recently stored in the brain can be integrated with other information and transformed into stable representations through a process known as systems-level consolidation (15). The mechanisms of this transformation are thought to involve repeated reactivation of information, particularly during sleep, leading to subsequent improvement in postsleep memory performance (19–24).

Taking into consideration the role of sleep in memory consolidation, we adapted procedures for (i) reducing implicit social biases and (ii) reactivating this training during sleep. We were particularly interested in factors that can influence whether such training procedures produce transient or persistent effects. Because pervasive stereotypes in the media and broader culture could function to regenerate a bias that is momentarily reduced (14), maintaining the benefits of training is crucial for the ultimate usefulness of potential bias-reducing interventions.

We reactivated counterbias information during sleep using subtle auditory cues that had been associated with counterbias training. Participants were White males and females from a university community (N = 40) and were recruited as two subsamples that allowed for a direct replication (25). First, biases were quantified using two versions of the implicit association test (IAT) (26). The IAT allows for an assessment of the strength of implicit associations between social groups and attributes (26). One test examined the degree to which female faces were preferentially associated with art versus science words—or the re-

verse for male faces (gender-bias IAT). The other test examined the degree to which Black faces were preferentially associated with bad versus good words—or the reverse for White faces (racial-bias IAT). Results were quantified by using a conventional scoring procedure (27), in which zero indicates no bias and larger scores indicate greater bias. Consistent with previous research (7), IAT scores showed that participants held implicit social biases for both gender and race, with both scores significantly greater than zero [mean ± SEM, 0.559 ± 0.044; gender $t(39)$ = 9.076, $P <$ 0.001; race $t(39)$ = 8.388, $P < 0.001$].

After this confirmation of baseline levels of implicit bias, participants engaged in training designed to reduce gender and racial bias (13). In

both cases, bias reduction was expected because participants intentionally selected counterstereotype information intermixed with other information. Participants viewed several types of face-word pairing but were required to attend and respond only to pairings that countered the typical bias (Fig. 1A) (25). Two unusual frequency-modulated sounds were presented during training, one after correct counter–gender bias responses and the other after correct counter–racial bias responses. To reinforce these associations, we administered another task wherein the same two sounds prompted participants to form a corresponding face-word pairing (25). Training thus established a strong association between each sound and one type of counterbias training.
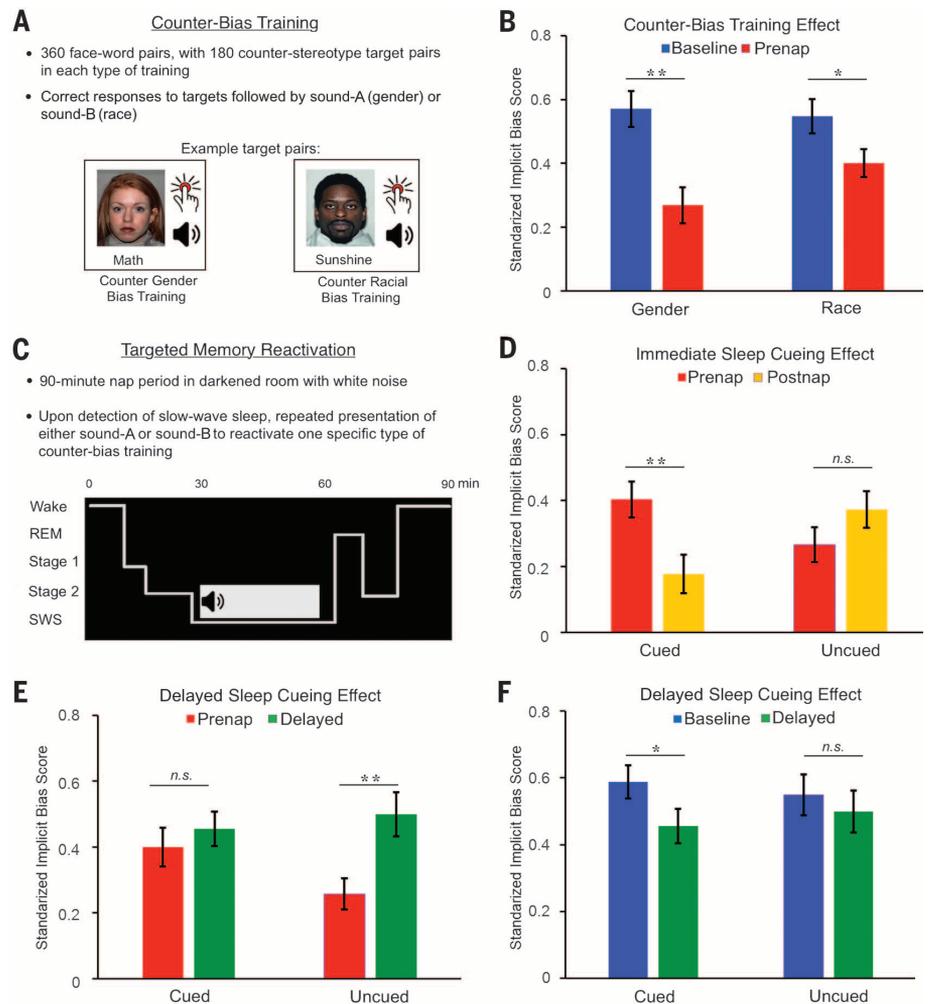
Fig. 1. Experimental procedures and results. (**A**) Procedures for counterbias training with sound cues. (**B**) Implicit bias reduction was found for both counter–racial bias training and counter–gender stereotype training. Bias was measured using the IAT before training (baseline) and after training (prenap). Error bars indicate ±1 SEM adjusted for within-subject comparisons. (**C**) Procedures for the nap phase of the experiment, when one sound was repeatedly played to participants during SWS, using a low intensity to avoid arousal from sleep. (**D**) The change in implicit bias from prenap to postnap diverged as a function of cueing condition, with a further reduction only for the cued social bias. (**E**) The change in implicit bias from prenap to the 1-week delay diverged as a function of cueing condition, with a significant increase only for the uncued social bias. (**F**) The change in implicit bias from baseline to the 1-week delay diverged as a function of cueing condition, with a significant reduction only for the cued social bias. Significant pairwise differences are indicated: *$P < 0.05$ or **$P ≤ 0.01$.

Biases were reduced compared with baseline levels (Fig. 1B) [within-subject analysis of variance (ANOVA), $F_{1,39}$ = 15.453, P < 0.001, η_p^2 = 0.284]. The mean IAT score was 0.559 at baseline and 0.335 at the prenap test. This bias reduction did not differ as a function of bias type ($F_{1,39}$ = 1.840, P = 0.183).

Next, participants were invited to take a 90-min afternoon nap (Fig. 1C; see table S1 for sleep-stage information). When electroencephalographic signals showed clear signs of slow-wave sleep (SWS), we repeatedly played one auditory cue, randomly selected as the counter–gender bias sound (n = 21) or the counter–racial bias sound (n = 19). Stimulation was discontinued at any sign of arousal from sleep. The number of presentations averaged 258 ± 24 (SEM).

Implicit biases were measured again after waking. Bias change from prenap to postnap varied with cueing condition as predicted [substantiated by a two-way interaction (cued or uncued by prenap or postnap), $F_{1,39}$ = 14.612, P < 0.001, η_p^2 = 0.273]. As shown in Fig. 1D, implicit bias was significantly reduced from prenap to postnap when cued [$t(39)$ = 2.698, P = 0.010] and unchanged when not cued [$t(39)$ = –1.378, P = 0.176] (fig. S1 and S2). This differential bias reduction was not moderated by bias type (fig. S3). Thus, reactivating counterbias learning during sleep can selectively reduce implicit racial or gender bias, depending on which form of counterbias training was cued.

Implicit biases were measured again after 1 week, revealing that the differential bias reduction endured (Fig. 1E) (n = 38; $F_{1,37}$ = 4.672, P = 0.037, η_p^2 = 0.112). Cueing during sleep resulted in sustained counterbias reduction, such that the cued bias did not differ between prenap and delayed testing

Fig. 2. Relation between long-term bias reduction and sleep physiology. The quality of sleep after training as indexed by the product of (minutes in SWS) × (minutes in REM sleep), predicted differential bias change, quantified as follows. Given that standardized implicit bias scores were preferentially reduced for the cued relative to uncued condition overall (Fig. 1F), we computed the reduction separately for cued and uncued conditions [(baseline score) − (delayed score) in both cases]. Differential bias change was taken as the cued reduction minus the uncued reduction, such that higher values indicated larger bias reduction over this interval for the cued compared to the uncued bias.

[$t(37)$ = −0.774, P = 0.444], whereas the uncued bias increased during the delay [$t(37)$ = –3.078, P = 0.004]. When compared with baseline (Fig. 1F), cued biases were weaker after 1 week [$t(37)$ = 2.203, P = 0.034], whereas uncued biases were not [$t(37)$ = 0.524, P = 0.603], although the interaction was not significant ($F_{1,37}$ = 0.471, P = 0.497).

Neurophysiological activity during sleep—such as sleep spindles, slow waves, and rapid-eye-movement (REM) duration—can predict later memory performance (*17*). Accordingly, we explored possible relations between cueing-specific bias reduction and measures of sleep physiology. We found that only SWS × REM sleep duration consistently predicted cueing-specific bias reduction at 1 week relative to baseline (Fig. 2) [$r(38)$ = 0.450, P = 0.005] (*25*).

Past research indicates that by pairing learning episodes with auditory or olfactory stimuli and then presenting these stimuli again during postlearning SWS, learned information can be specifically reactivated and strengthened (*19*). Benefits of this targeted memory reactivation (TMR) have been documented for declarative, procedural, and emotional memories (*19*). Such learning typically does not challenge preexisting knowledge nor compete with daily experiences outside the laboratory. In contrast, we examined learning-induced changes in long-standing social biases. We showed that selectively reactivating counterbias learning during sleep weakened preexisting implicit social biases immediately after the nap and facilitated the retention of this learning going forward. Without TMR during sleep, training effects tended to dissipate, and the bias returned to baseline levels. These results thus enlarge our conception of sleep's role in socially relevant learning.

Observed relations between sleep neurophysiology and behavior further reinforced the conclusion that bias reduction is sleep-dependent. Current thinking about consolidation emphasizes sets of cortical networks that can become integrated through interactions with hippocampal networks, possibly by means of cyclic SWS-REM periods (*15*–*17*, *28*). The correlation with SWS × REM duration implicates a benefit from REM-based processing subsequent to SWS-based reactivation, perhaps to integrate learning within associative knowledge networks. These findings support the notion that both SWS and REM are operative in sleep-dependent memory consolidation (*16*, *17*, *28*, *29*).

Future research is needed to address many outstanding questions in relation to our findings. For example, how much training is needed to make implicit benefits persist for long periods of time and transfer to explicit benefits in interpersonal interactions? To what extent do persistent benefits depend on repeated training, the nature of other waking activities after training, and repeated memory reactivation during sleep? Although IAT measures are imperfect and may sometimes reflect knowledge of cultural stereotypes rather than implicit bias per se (*30*), prior research has demonstrated consequences for social behavior, such that low implicit bias as measured with the IAT may

indeed be linked with egalitarianism (*6*, *7*). Given that training to reduce implicit bias can be conceptualized as a type of habit learning (*31*), perhaps novel sleep manipulations could be adapted to aid people in changing various unwanted or maladaptive habits, such as smoking, unhealthy eating, catastrophizing, or selfishness (*32*).

REFERENCES AND NOTES

1. J. Glaser, *Suspect Race: Causes and Consequences of Racial Profiling* (Oxford Univ. Press, New York, 2014).
2. J. Correll *et al.*, *J. Pers. Soc. Psychol.* **92**, 1006–1023 (2007).
3. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474–16479 (2012).
4. L. Bobo, C. L. Zubrinsky, *Soc. Forces* **74**, 883–909 (1996).
5. P. G. Devine, *J. Pers. Soc. Psychol.* **56**, 5–18 (1989).
6. D. A. Stanley, P. Sokol-Hessner, M. R. Banaji, E. A. Phelps, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7710–7715 (2011).
7. A. G. Greenwald, T. A. Poehlman, E. L. Uhlmann, M. R. Banaji, *J. Pers. Soc. Psychol.* **97**, 17–41 (2009).
8. B. A. Nosek *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10593–10597 (2009).
9. G. W. Allport, *The Nature of Prejudice* (Addison-Wesley, Reading, MA, 1979).
10. S. Fiske, in *The Handbook of Social Psychology,* D.T. Gilbert, S.T. Fiske, G. Lindzey, Eds. (Oxford Univ. Press, New York, ed. 4, 1998); vol. 2, pp. 357–411.
11. A. Olsson, J. P. Ebert, M. R. Banaji, E. A. Phelps, *Science* **309**, 785–787 (2005).
12. C. S. Crandall, A. Eshleman, L. O'Brien, *J. Pers. Soc. Psychol.* **82**, 359–378 (2002).
13. B. Gawronski, R. Deutsch, S. Mbirkou, B. Seibt, F. Strack, *J. Exp. Soc. Psychol.* **44**, 370–377 (2008).
14. M. Weisbuch, K. Pauker, N. Ambady, *Science* **326**, 1711–1714 (2009).
15. K. A. Paller, in *Encyclopedia of Neuroscience*, L. R. Squire, Ed. (Academic Press, Oxford, 2009), pp. 741–749.
16. B. Rasch, J. Born, *Physiol. Rev.* **93**, 681–766 (2013).
17. S. Diekelmann, J. Born, *Nat. Rev. Neurosci.* **11**, 114–126 (2010).
18. R. Stickgold, M. P. Walker, *Nat. Neurosci.* **16**, 139–145 (2013).
19. D. Oudiette, K. A. Paller, *Trends Cogn. Sci.* **17**, 142–149 (2013).
20. B. Rasch, C. Büchel, S. Gais, J. Born, *Science* **315**, 1426–1429 (2007).
21. J. W. Antony, E. W. Gobel, J. K. O'Hare, P. J. Reber, K. A. Paller, *Nat. Neurosci.* **15**, 1114–1116 (2012).
22. M. A. Wilson, B. L. McNaughton, *Science* **265**, 676–679 (1994).
23. P. Peigneux *et al.*, *Neuron* **44**, 535–545 (2004).
24. J. D. Rudoy, J. L. Voss, C. E. Westerberg, K. A. Paller, *Science* **326**, 1079 (2009).
25. Materials and methods are available as supplementary materials on *Science* online.
26. A. G. Greenwald, D. E. McGhee, J. L. K. Schwartz, *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
27. A. G. Greenwald, B. A. Nosek, M. R. Banaji, *J. Pers. Soc. Psychol.* **85**, 197–216 (2003).
28. R. Stickgold, D. Whidbee, B. Schirmer, V. Patel, J. A. Hobson, *J. Cogn. Neurosci.* **12**, 246–254 (2000).
29. M. V. Ambrosini, A. Giuditta, *Sleep Med. Rev.* **5**, 477–490 (2001).
30. H. R. Arkes, P. E. Tetlock, *Psychol. Inq.* **15**, 257–278 (2004).
31. P. G. Devine, P. S. Forscher, A. J. Austin, W. T. L. Cox, *J. Exp. Soc. Psychol.* **48**, 1267–1278 (2012).
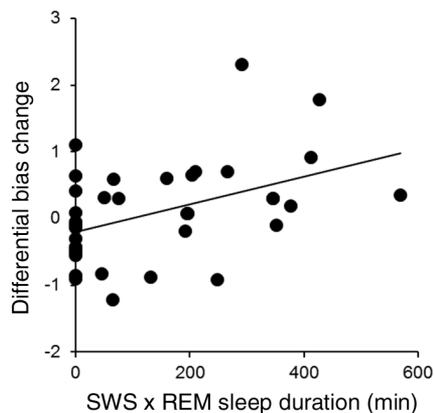32. A. Arzi *et al.*, *J. Neurosci.* **34**, 15382–15393 (2014).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6238/1013/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
Sound Files
References (*33*–*36*)

Science

Unlearning implicit social biases during sleep

Xiaoqing Hu, James W. Antony, Jessica D. Creery, Iliana M. Vargas, Galen V. Bodenhausen and Ken A. Paller

Sleep on it: Consolidating implicit learning

A good night's sleep is one of the best ways to fix recently learned information into long-lasting memory. Recent evidence suggests that recent memories are reactivated during sleep and woven into existing representations of stored information. Hu *et al.* now demonstrate that triggering memory consolidation during sleep can help set into place recently learned anti-bias training (see the Perspective by Feld and Born). Changes in people's stereotypical attitudes toward race and gender were maintained for up to 1 week after training.

Science, this issue p. 1013; see also p. 971

ARTICLE TOOLS	http://science.sciencemag.org/content/348/6238/1013
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2015/05/27/348.6238.1013.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/348/6238/971.full
REFERENCES	This article cites 29 articles, 9 of which you can access for free http://science.sciencemag.org/content/348/6238/1013#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service