Science AAAS

www.sciencemag.org/content/348/6238/1013/suppl/DC1

Supplementary Materials for

Unlearning implicit social biases during sleep

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

*Corresponding author. E-mail: kap@northwestern.edu

Published 29 May 2015, *Science* **348**, 1013 (2015)
DOI: 10.1126/science.aaa3841

This PDF file includes

Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/348/6238/1013/suppl/DC1)

Sound files S1 and S2

Materials and Methods

Participants: Forty White participants (20 females, age: $M \pm SD$, 21.75 ± 3.60 years old, range: 19–32 years old) were recruited from the local university community. The study was approved by the Northwestern University Institutional Review Board and all participants gave informed consent. Data from 17 additional participants were excluded from further analyses because they did not enter slow-wave sleep (n = 12), reported hearing sound cues during sleep (n = 4), or an EEG data file was corrupted (n = 1). Data were collected from two independent samples, allowing for a direct replication. Data from sample-1 (n = 21) were collected during June–July 2014; data from sample-2 (n = 19) were collected during February–March 2015. On the day before the experiment, all participants were reminded via email that they should wake up at least 2 hours earlier than usual and refrain from caffeine the next day.

Stimuli: Neutral face stimuli (Black and White, both genders) were chosen from the Eberhardt Lab Face Database, the NimStim Face Stimulus Set, and the Karolinska Directed Emotional Faces (*32, 33*). Word stimuli (categorized as good, bad, science, or art) were chosen from the literature on implicit racial bias and gender stereotypes (*8, 25, 34*). Different stimulus sets were used in the baseline administration of Implicit Attitude Tests (IATs), the counterbias training, the sound cue–retrieval task, and the subsequent IATs. Words used in both the test and the training sessions are provided below:

Good/bad words used in the racial-bias IATs: *Paradise, Freedom, Harmony, Smile, Cheer, Sunrise, Disaster, Vomit, Bomb, Cancer, Rotten, Coma*

Good/bad words used in the counter–racial bias training: *Heaven, Pleasure, Sunshine, Miracle, Honor, Evil, Murder, Grief, Poison, Virus,*

Science/art words used in the gender-bias IATs: *Geology, Physics, Einstein, Experiment, NASA, Biology, Poetry, Literature, Shakespeare, Symphony, English, Humanities*

Science/art words used in the counter–gender bias training: *Science, Chemistry, Math, Geometry, Engineering, Dance, Theater, Drama, History, Music*

Procedure: All participants completed the following procedures: (i) baseline IATs, ~10 min; (ii) counterbias training sessions with sound cues, ~30 min; (iii) EEG set-up, ~20 min; (iv) prenap IATs, ~10 min; (v) sound cue–retrieval task, ~5 min; (vi) 90-min nap plus a 10-min break; (vii) postnap IATs, ~10 min; and (viii) 1-week delayed IATs, ~10 min. For the delayed tests, 38 out of 40 participants returned. Among these, 31 participants returned after 7 days, and the remaining returned either 6 (n = 3), 8 (n = 2), 9 (n = 1), or 10 days later (n = 1). All tests were conducted between 10 a.m. and 6 p.m., with the nap starting between 12:30 p.m. and 4:30 p.m. Experimenters were blind to the cueing conditions during IAT administrations. Speed and accuracy were emphasized equally during all tasks.

Baseline Implicit Bias Assessments. Participants completed two IATs, a racial-bias IAT and a gender-stereotype IAT. Both IATs were structured in the seven-block version [see reference (*26*) for task structure details]. Twenty-one participants completed the racial IAT first and 19 completed the gender IAT first. In the racial IAT, participants performed a categorization task that included good and bad words randomly mixed with White and Black male faces. In the gender-stereotype IAT, participants performed a similar task but with science and art words randomly mixed with White female and male faces. On each trial, a single stimulus was presented centrally until a correct response was registered (onset 150 ms after prior response). If an incorrect response was registered, an error feedback "X" was presented on the screen until participants gave the correct response. Participants were instructed to make a categorization response as quickly and accurately as possible. There were always two response buttons (buttons "E" and "I" on a keyboard), but button assignments alternated across assessment blocks (each with 60 trials). Assignments were either consistent or inconsistent with the typical bias. In the former case, one button was used for White+good (or male+science) and the other for Black+bad (or female+art). In the latter case, one button was used for Black+good (or female+science) and the other for White+bad (or male+art). Implicit racial or gender bias was inferred when categorization responses were quicker and more accurate in blocks with button assignments that were consistent with the typical bias compared with blocks with button assignments that were inconsistent with the typical bias.

To quantify IAT results, a D_{600} score was calculated as the dependent variable following a conventional algorithm (*26*). First, response times (RTs) shorter than 300 ms or longer than 3 s were deleted (<1%). Second, RTs for correct responses were averaged separately for blocks with button assignments consistent with the typical bias and for blocks with button assignments inconsistent with the typical bias. Third, we calculated the standard deviation of the RT distributions from correct trials of both consistent and inconsistent blocks combined. Fourth, any incorrect responses were replaced with the mean RT associated with that particular block plus a 600-ms penalty (*26*). Fifth, the means of consistent and inconsistent blocks were calculated separately including RTs of incorrect responses with the error penalties. Sixth, the RT differences between the consistent and the inconsistent blocks ($RT_{inconsistent\ -\ consistent}$) from step five were divided by the inclusive standard deviation obtained from step three. The result of step six was the D_{600} score (*26*). A larger D_{600} score indicates a stronger implicit social bias. The measure does not indicate whether there is also an explicit social bias, or whether there are interactions between processing related to implicit and explicit measures of social bias.

Counterbias Training with Sound Cues: Participants completed a counter–racial bias training session and a counter–gender bias training session, and the order of the two sessions was counterbalanced (counter–racial bias training first for 20 participants and counter–gender bias training first for 20 participants). During each training session, participants were presented with 360 trials (divided into 3 blocks each followed by a short break, and with trials separated by an interstimulus interval of 1 s). Half of the trials were counterbias face+word pairs: Black+good or female+science. On each of these trials, the participant was required to press the spacebar. Critically, participants were informed that if they made a quick and accurate response to a counterbias pair, correct feedback would be given in the form of a 1-s sound from a speaker (for this purpose, quick responses were those made within 800 ms, though participants were not directly informed of this specific response deadline). One of the two training sounds, sound-a or sound-b, (Sounds S1 and S2, respectively) was used for this feedback in each counterbias training session (counterbalanced across subjects; sound-a was used in counter–racial bias training for half of the participants, and in counter–gender bias training for the remaining participants; sounds and other stimuli are available from the first author upon request). The remaining trials were fillers that required no response. Filler trials included 60 trials in each of three categories: Black+bad; White+good; and White+bad (counter–racial bias training) or female+art; male+science; and male+art (counter–gender bias training). Filler trials were included so that participants actively discriminated between counterbias and filler trials, with an emphasis on selecting counterbias trials.

EEG Setup. After counterbias training, an EEG cap with 21 electrodes was applied. EOG and EMG electrodes were also attached to monitor eye movement and muscle activity. Continuous EEG was recorded from International 10–20 locations Fpz, Fz, Cz, Pz, Oz, Fp1/2, F3/4, F7/8, C3/4, P3/4, T3/4, T5/6, and O1/2. Two electrodes were placed to monitor eye movements, one below the left eye and the other next to the right eye for recording the vertical and horizontal EOG; one electrode was placed on the chin to record EMG. Online EEG recordings were referenced to the average of left and right mastoid, filtered at 0.1–200 Hz with a 60 Hz notch filter, and sampled at 500 Hz.

Prenap Implicit Bias Assessments: Participants were tested with the two IATs again to measure training effects. The order of the two tests was the same as in the baseline measurements, but the word and face stimuli used in each IAT were unique.

Sound-Cue-Retrieval Task: This task included 120 trials that were divided into six blocks, with trials separated by an interstimulus interval of 1 s. Each trial began with presentation of one of the two sounds that were used in the counterbias training tasks. Concurrently with the onset of the sound, either a White female face or a Black male face was presented on the left side of the monitor, while a science word and a good word were presented on the upper-right or lower-right side of the monitor (location of words randomized across trials). Participants were required to use a mouse to drag the face to its corresponding word to create a counterbias pair in accordance with the sound that was presented. They were told that if they heard the sound from counter–gender bias training, they were to link the female with the science word, and if they heard the sound from counter–racial bias training, they were to link the Black face with the good word. This task thus provided further training in associating each sound with its corresponding counterbias associations, as well as active training in counterbias face-word associations. It thus ensured that the sound-to-bias–type associations were well learned. It is also notable that the sounds were presented as a signal to retrieve counterbias information in this task, whereas during counterbias training the sounds were presented only after correct responses. We cannot determine which aspects of these procedures were most relevant, but the combination was apparently effective in strongly linking each sound to a specific type of counterbias training such that presenting sound cues during sleep could reactivate the corresponding memory.

Nap. Participants were prepared for a 90-min nap session in the same room, with a futon pillow, blanket, and sheets supplied. Participants were told that white noise would be played over a speaker to block out any environmental noise. When the experimenter observed that the participant had entered slow-wave sleep on the basis of standard EEG criteria, a sound cue was repeatedly presented; 19 participants received the sound associated with counter–racial bias training and 21 received the sound associated with counter–gender-stereotype training. Each sound cue lasted for 1 s, with an interstimulus interval of 4 s. The intensity of the cue was similar to the background white noise [38–40 db SPL]. Sound cues were halted whenever arousal was evidenced in physiological recordings. This sleep phase ended after 90 min unless the participant was still in slow-wave sleep (n = 3), in which case additional time was allowed for the slow-wave sleep stage to end naturally.

Sleep Physiological Analysis. Experimenters who were blind to participants' results conducted all sleep physiological analyses. Continuous EEG traces were down-sampled to 125 Hz and filtered at 0.5–60 Hz using a two-way least squares finite impulse response ("eeg_filt" function) offline. Sleep stages were formally identified offline using standard sleep-scoring methods (*35*).

Postnap Implicit Bias Assessments. This session began 10 min after the participant woke up from the nap. Participants finished two IATs, the order of which was opposite from baseline measurements. After the postnap bias assessments, participants were asked in an open manner whether they had heard any sound during sleep other than the white noise. Participants who reported that they heard a sound associated with counterbias training were excluded (n =4).

Delayed Implicit Bias Assessments. Participants returned to the lab for the two IATs after 1 week. Implicit biases were assessed using the same methods as used in the postnap implicit bias assessment.

Supplementary Text

Physiology-behavior correlations

We calculated cueing advantage effects in several ways. For cued and uncued biases separately, we focused on three comparisons: (i) prenap bias minus postnap bias to reflect the immediate bias change, (ii) prenap bias minus delayed bias to reflect the persistent bias change and (iii) baseline bias minus delayed bias to reflect the persistent bias change. We then subtracted the uncued bias change from the cued bias change for each of the three comparisons, which yielded a measure for Differential Bias Change. A positive value indicates that cued biases were reduced more than uncued biases (i.e., sleep-dependent cueing advantage). Results showed that only the baseline-versus-delayed Differential Bias Change values were positively correlated with the product of SWS and REM sleep durations [$r(38) = 0.450$, 95% confidence interval (CI): 0.212, 0.658; Fig. 2]. This correlation was highly consistent across sample-1 and sample-2 [$r(19) = 0.443$, 95% CI: –0.001, 0.775, and $r(19) = 0.568$, 95% CI: 0.220, 0.792, respectively]. The variability in scores across participants can also be viewed in a scatterplot of IAT results from each test separately (Fig. S1).

Subsamples

In both sample-1 and sample-2, we found an impact of memory reactivation during sleep on bias reduction. The primary comparison was for prenap versus postnap IAT scores as a function of sleep cueing condition (cued/uncued), as shown in Fig. S2. Indeed, cueing one type of counterbias training during sleep revealed a selective reduction in implicit bias from prenap to postnap, both in sample-1 [$F(1,20) = 7.027$, $P = 0.015$, $\eta_p^2 = 0.260$] and in sample-2 [$F(1,18) = 7.382$, $P = 0.014$, $\eta_p^2 = 0.291$].

Although prenap biases were not closely matched for cued and uncued conditions in sample-2, procedures for these two conditions were identical at this point. The failure to have a close match can thus be attributed to the randomization process for sample-2. However, we can be confident that the central sleep cueing interaction (Fig. S2; cued/uncued by prenap/postnap) is not dependent on prenap differences, because the same interaction was found in sample-1 when cued and uncued conditions were closely matched (Fig. S2A). Moreover, prenap differences in sample-2 were driven by lower uncued gender bias scores, as shown in Fig. S3.

Effects of bias type

To examine whether bias type (gender vs. race) modulated immediate or delayed sleep cueing effects, we conducted two mixed ANOVAs on IAT scores (Fig. S3). Both ANOVAs included bias type (gender/race) as a between-subject variable, with cueing (cued/uncued) and time (prenap/postnap in the first ANOVA, prenap/delayed in the second ANOVA) as within-subject variables. Results showed that the three-way interactions were not significant, suggesting that bias type did not modulate either the prenap/postnap sleep cueing effect [$F(1,38) = 0.212$, $P = 0.648$, $\eta_p^2 = 0.006$], or the prenap/delayed sleep cueing effect [$F(1,36) = 0.006$, $P = 0.939$, $\eta_p^2 = 0.001$].

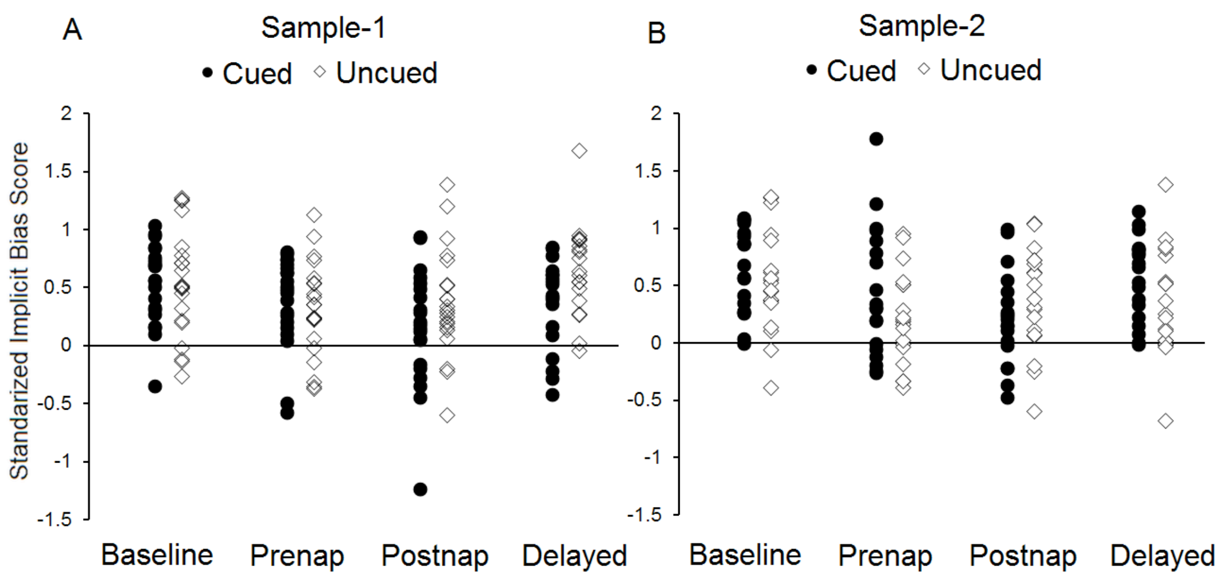

Fig. S1. Individual participant data for each sub-sample for baseline, prenap, postnap and 1-week delayed IAT assessments.

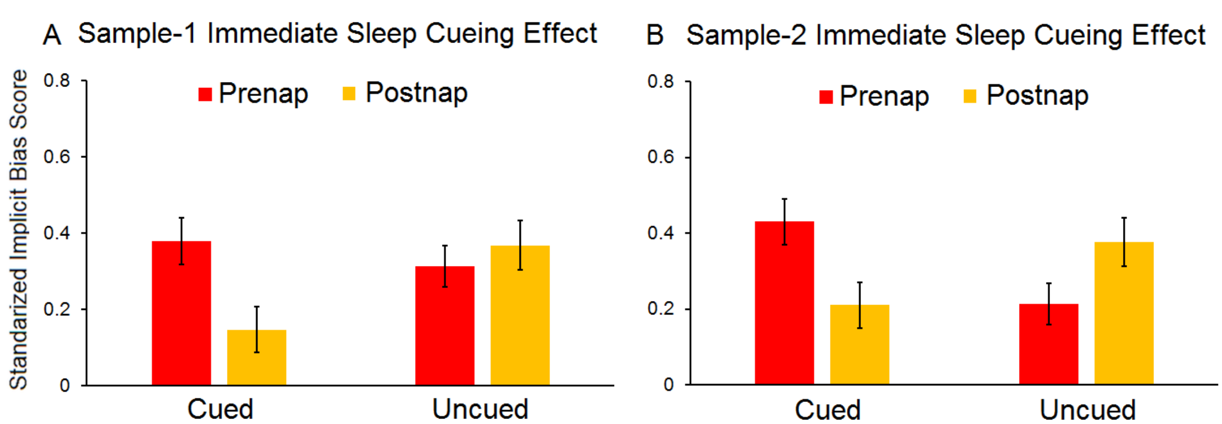

Fig. S2. Replication of the prenap versus postnap sleep cueing effect in sample-1 and sample-2. Cue presentations during sleep led to further bias reduction after a nap in both samples. In contrast, no reduction was observed for uncued biases. Error bars indicate ±1 SEM adjusted for within-subject comparisons.

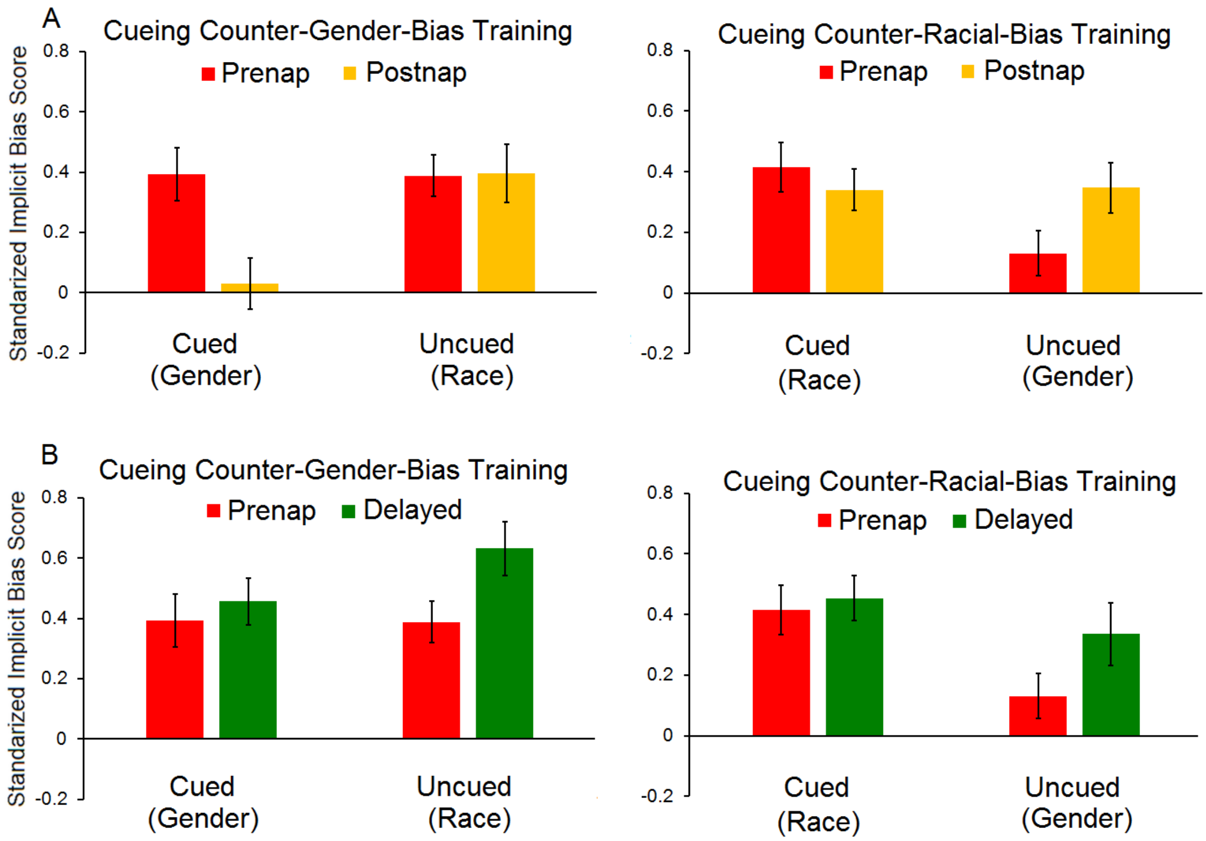

Fig. S3. Behavioral data analyzed separated by type of bias. When counter–gender bias training was cued, racial bias was uncued; when counter–racial bias training was cued, gender bias was uncued. (**A**) Similar patterns of prenap versus postnap sleep cueing effects were found in both cases. (**B**) Similar patterns of prenap versus delayed sleep cueing effects were found in both cases. Error bars indicate ±1 SEM adjusted for within-subject comparisons.

Table S1. Sleep stage results.

Sleep Stage	Time (Mean min ± SEM)
Wake	21.38 ± 3.11
Stage 1	7.33 ± 0.92
Stage 2	26.03 ± 1.67
Slow-Wave Sleep	27.16 ± 2.40
REM	5.35 ± 0.92

REFERENCES AND NOTE

1. J. Glaser, *Suspect Race: Causes and Consequences of Racial Profiling* (Oxford Univ. Press, New York, 2014).

2. J. Correll, B. Park, C. M. Judd, B. Wittenbrink, M. S. Sadler, T. Keesee, Across the thin blue line: Police officers and racial bias in the decision to shoot. *J. Pers. Soc. Psychol.* **92**, 1006–1023 (2007). Medline doi:10.1037/0022-3514.92.6.1006

3. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474–16479 (2012). Medline

4. L. Bobo, C. L. Zubrinsky, Attitudes on residential integration: Perceived status differences, mere in-group preference, or racial prejudice? *Soc. Forces* **74**, 883–909 (1996). doi:10.1093/sf/74.3.883

5. P. G. Devine, Stereotypes and prejudice: Their automatic and controlled components. *J. Pers. Soc. Psychol.* **56**, 5–18 (1989). doi:10.1037/0022-3514.56.1.5

6. D. A. Stanley, P. Sokol-Hessner, M. R. Banaji, E. A. Phelps, Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7710–7715 (2011). Medline doi:10.1073/pnas.1014345108

7. A. G. Greenwald, T. A. Poehlman, E. L. Uhlmann, M. R. Banaji, Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* **97**, 17–41 (2009). Medline doi:10.1037/a0015575

8. B. A. Nosek, F. L. Smyth, N. Sriram, N. M. Lindner, T. Devos, A. Ayala, Y. Bar-Anan, R. Bergh, H. Cai, K. Gonsalkorale, S. Kesebir, N. Maliszewski, F. Neto, E. Olli, J. Park, K. Schnabel, K. Shiomura, B. T. Tulbure, R. W. Wiers, M. Somogyi, N. Akrami, B. Ekehammar, M. Vianello, M. R. Banaji, A. G. Greenwald, National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10593–10597 (2009). Medline doi:10.1073/pnas.0809921106

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, The role of social groups in the persistence of learned fear. *Science* **309**, 785–787 (2005). Medline doi:10.1126/science.1113551

12. C. S. Crandall, A. Eshleman, L. O'Brien, Social norms and the expression and suppression of prejudice: The struggle for internalization. *J. Pers. Soc. Psychol.* **82**, 359–378 (2002). Medline doi:10.1037/0022-3514.82.3.359

13. B. Gawronski, R. Deutsch, S. Mbirkou, B. Seibt, F. Strack, When "Just Say No" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *J. Exp. Soc. Psychol.* **44**, 370–377 (2008). doi:10.1016/j.jesp.2006.12.004

14. M. Weisbuch, K. Pauker, N. Ambady, The subtle transmission of race bias via televised nonverbal behavior. *Science* **326**, 1711–1714 (2009). Medline doi:10.1126/science.1178358

15. K. A. Paller, in *Encyclopedia of Neuroscience*, L. R. Squire, Ed. (Academic Press, Oxford, 2009), pp. 741–749.

16. B. Rasch, J. Born, About sleep's role in memory. *Physiol. Rev.* **93**, 681–766 (2013). Medline doi:10.1152/physrev.00032.2012

17. S. Diekelmann, J. Born, The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010). Medline

18. R. Stickgold, M. P. Walker, Sleep-dependent memory triage: Evolving generalization through selective processing. *Nat. Neurosci.* **16**, 139–145 (2013). Medline doi:10.1038/nn.3303

19. D. Oudiette, K. A. Paller, Upgrading the sleeping brain with targeted memory reactivation. *Trends Cogn. Sci.* **17**, 142–149 (2013). Medline doi:10.1016/j.tics.2013.01.006

20. B. Rasch, C. Büchel, S. Gais, J. Born, Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* **315**, 1426–1429 (2007). Medline doi:10.1126/science.1138581

21. J. W. Antony, E. W. Gobel, J. K. O'Hare, P. J. Reber, K. A. Paller, Cued memory reactivation during sleep influences skill learning. *Nat. Neurosci.* **15**, 1114–1116 (2012). Medline doi:10.1038/nn.3152

22. M. A. Wilson, B. L. McNaughton, Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994). Medline doi:10.1126/science.8036517

23. P. Peigneux, S. Laureys, S. Fuchs, F. Collette, F. Perrin, J. Reggers, C. Phillips, C. Degueldre, G. Del Fiore, J. Aerts, A. Luxen, P. Maquet, Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* **44**, 535–545 (2004). Medline doi:10.1016/j.neuron.2004.10.007

24. J. D. Rudoy, J. L. Voss, C. E. Westerberg, K. A. Paller, Strengthening individual memories by reactivating them during sleep. *Science* **326**, 1079 (2009). Medline doi:10.1126/science.1179013

25. Materials and methods are available as supplementary materials on *Science* online.

26. A. G. Greenwald, D. E. McGhee, J. L. K. Schwartz, Measuring individual differences in implicit cognition: The implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998). Medline doi:10.1037/0022-3514.74.6.1464

27. A. G. Greenwald, B. A. Nosek, M. R. Banaji, Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197–216 (2003). Medline doi:10.1037/0022-3514.85.2.197

28. R. Stickgold, D. Whidbee, B. Schirmer, V. Patel, J. A. Hobson, Visual discrimination task improvement: A multi-step process occurring during sleep. *J. Cogn. Neurosci.* **12**, 246–254 (2000). Medline doi:10.1162/089892900562075

29. M. V. Ambrosini, A. Giuditta, Learning and sleep: The sequential hypothesis. *Sleep Med. Rev.* **5**, 477–490 (2001). Medline doi:10.1053/smrv.2001.0180

30. H. R. Arkes, P. E. Tetlock, Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test?”. *Psychol. Inq.* **15**, 257–278 (2004). doi:10.1207/s15327965pli1504_01

31. P. G. Devine, P. S. Forscher, A. J. Austin, W. T. L. Cox, Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *J. Exp. Soc. Psychol.* **48**, 1267–1278 (2012). Medline doi:10.1016/j.jesp.2012.06.003

32. A. Arzi, Y. Holtzman, P. Samnon, N. Eshel, E. Harel, N. Sobel, Olfactory aversive conditioning during sleep reduces cigarette-smoking behavior. *J. Neurosci.* **34**, 15382–15393 (2014). Medline doi:10.1523/JNEUROSCI.2291-14.2014

33. D. Lundqvist, A. Flykt, A. Öhman, *The Karolinska Directed Emotional Faces— KDEF* [CD-ROM] (Psychology Section, Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden, 1998).

34. N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, C. Nelson, The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res.* **168**, 242–249 (2009). Medline doi:10.1016/j.psychres.2008.05.006

35. B. A. Nosek, M. R. Banaji, A. G. Greenwald, Math = male, me = female, therefore math not = me. *J. Pers. Soc. Psychol.* **83**, 44–59 (2002). Medline doi:10.1037/0022-3514.83.1.44

36. C. Iber, S. Ancoli-Israel, A. Chesson, S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications* (American Academy of Sleep Medicine, Westchester, IL, ed. 1, 2007).