Check for updates

Research Report

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
2015, Vol. 26(7) 1098–1106

DOI: 10.1177/0956797615575734
pss.sagepub.com
SAGE

Suppressing Unwanted Autobiographical Memories Reduces Their Automatic Influences: Evidence From Electrophysiology and an Implicit Autobiographical Memory Test

Xiaoqing Hu[1,2], Zara M. Bergström[3], Galen V. Bodenhausen[1], and J. Peter Rosenfeld[1]
[1]Northwestern University, [2]University of Texas at Austin, and [3]University of Kent

Abstract

The present study investigated the extent to which people can suppress unwanted autobiographical memories in a memory-detection context involving a mock crime. Participants encoded sensorimotor-rich memories by enacting a lab-based crime (stealing a ring) and received instructions to suppress memory of the crime in order to evade guilt detection in a brain-wave-based concealed-information test. Aftereffects of suppression on automatic memory processes were measured in an autobiographical Implicit Association Test. Results showed that suppression attenuated brain-wave activity (the P300) associated with crime-relevant memory retrieval, which rendered waveforms from innocent and guilty participants indistinguishable. However, the two groups could nevertheless be discriminated via the late-posterior-negative slow wave, which may reflect the need to monitor response conflict arising between voluntary suppression and automatic recognition processes. Finally, extending recent findings that suppression can impair implicit memory processes, we provide novel evidence that suppression reduces automatic cognitive biases often associated with actual autobiographical memories.

Keywords
memory suppression, memory detection, autobiographical memory, P300, autobiographical Implicit Association Test, neuroscience and law, open data, open materials

Received 5/20/14; Revision accepted 2/10/15

The automatic intrusion of unwelcome memories can sting. People commonly rely on inhibitory control to prevent unwanted memories from intruding into awareness, which reduces explicit recall of such memories (Anderson & Green, 2001). Neuroimaging research suggests that suppressing previously encoded words and pictures involves mechanisms of cognitive control in the prefrontal cortex that down-regulate retrieval-related neural circuits in the hippocampus (Anderson & Hanslmayr, 2014; Depue, 2012). However, research has not yet examined suppression of autobiographical memories that people spontaneously desire to control in everyday life, such as memories of personal acts associated with guilt or shame. Thus, it is unknown whether people can directly suppress brain activity associated with sensorimotor-rich memories arising from autobiographical experiences and whether suppressed autobiographical memories are nevertheless implicitly active. Answering these questions could illuminate theoretical issues in cognitive control, as well as offer practical applications regarding neuroscientific approaches to guilt detection in translational fields such as neurolaw (Farah, Hutchinson, Phelps, & Wagner, 2014).

Corresponding Author:
Xiaoqing Hu, University of Texas at Austin, Psychology Department, 108 E. Dean Keeton Stop A8000, Austin, TX 78712
E-mail: xqhu@utexas.edu

We investigated these issues in a memory-detection context. Participants were asked to suppress sensorimotor-rich memories that were encoded during a lab-based crime. We hypothesized that suppressing autobiographical memory can attenuate the P300, an event-related potential (ERP) indicating conscious recollection (Paller, Kutas, & McIsaac, 1995; Rugg & Curran, 2007; Vilberg, Moosavi, & Rugg, 2006) that has long been used in memory detection (Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013). Indeed, retrieval suppression can reduce the amplitude of P300s in response to previously learned words (Bergström, de Fockert, & Richardson-Klavehn, 2009; Depue et al., 2013) and pictures in memory-detection tests (Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013).

We then measured how suppression modulated automatic influences of autobiographical memory in an autobiographical Implicit Association Test (aIAT), which uses simple cognitive judgments to assess whether autobiographical statements are automatically associated with truthfulness. Specifically, participants read statements that could describe a past autobiographical activity (e.g., "I took a ring") and had to classify these statements in terms of their general topic (as a ring-related event or not). On intermixed trials, they were asked to confirm or deny unequivocally true statements (e.g., "I am sitting in front of a computer") or false statements (e.g., "I am climbing a mountain"). The veracity of the autobiographical statements can be inferred from the speed and accuracy with which these simple classifications are made (Agosta & Sartori, 2013).

Notably, even if explicit memory retrieval is impaired by suppression, automatic memory processes may nevertheless remain intact, a well-documented dissociation (Schacter, 1987). Alternatively, top-down suppression can weaken memories' intrusions into awareness and also their automatic influences (Benoit, Hulbert, Huddleston, & Anderson, 2015; Levy & Anderson, 2012). Recent research shows that suppressing perceptual memories impaired object identifications in perceptual priming tasks (Gagnepain, Henson, & Anderson, 2014; Kim & Yi, 2013). We thus hypothesized that suppression can even weaken the automatic influence of sensorimotor-rich, autobiographical memories.

Method

Participants

We decided a priori on a sample size of 26 participants per group because a power analysis indicated that this number of participants was required to detect a large suppression effect (Cohen's $d = 0.8$) with a power of 0.8 at an alpha of .05; we expected a large effect in suppressing incidentally encoded crime-relevant memories given that (a) a recent meta-analysis in memory detection suggests that the P300 is extremely sensitive to variations of recognition (Meijer, Selle, Elber, & Ben-Shakhar, 2014) and (b) the most relevant prior memory-suppression research has typically produced medium to large suppression effects (Bergström et al., 2013; Gagnepain et al., 2014; Kim & Yi, 2013; Noreen & MacLeod, 2013). This sample size is also consistent with those used in relevant prior memory-suppression studies (which typically involved 24 participants per experiment or condition; e.g., Bergström et al., 2013; Gagnepain et al., 2014; Kim & Yi, 2013).

Seventy-eight participants from three experimental groups were included in the final analyses; 24 additional participants were excluded either because they had electroencephalographic (EEG) artifacts (n = 15) or because they did not follow instructions (n = 9; see the Supplemental Material available online). Participants were compensated with either course credit or money and were additionally promised a $10 reward if an innocent outcome was obtained from the brain-wave-based test. They were later given this $10 regardless of their performance. The study was approved by the Northwestern University Institutional Review Board.

Procedure

Each participant was randomly assigned to one of three groups (n = 26 per group). The standard-guilt group received no memory-suppression instructions, the suppressed-guilt group was given memory-suppression instructions, and the innocent group (which was not asked to commit the lab-based crime) was not given suppression instructions. Except as noted, all participants enacted either a lab-based crime or an innocent act (~10 min) and then completed an ERP-based concealed-information test (CIT; ~30 min) and an aIAT (~10 min). Participants in the two guilty groups also completed post-experiment questionnaires (~3 min).

The lab-based crime that participants in the two guilty groups were instructed to enact consisted of finding and stealing an object (a ring) from a faculty member's mailbox in the Psychology Department office, which is off limits to students. The word "ring" was never mentioned in the instructions. Thus, participants acquired the crime-relevant memory solely from enacting the crime. Innocent participants were instructed to go to the same area but to simply write their initials on a poster board near the office. They were thus unaware of any lab-based crime.

Next, all participants completed the CIT while continuous EEGs were recorded. Participants in the suppressed-guilt group received direct suppression instructions beforehand (Benoit & Anderson, 2012; Bergström et al., 2009): They were told they should never allow the memory of the lab-based crime to come to mind at all during the test, and they should not engage in distracting thoughts (see the Supplemental Material). Participants in the other two groups were not given any suppression instructions.

We employed the complex trial version of the CIT (see the Supplemental Material), which is more resistant to countermeasures than other CIT versions (Rosenfeld et al., 2013). On each trial, participants were presented with one of the following items for 300 ms: a probe (the word "ring") or one of six irrelevant stimuli (other words: "bracelet," "necklace," "watch," "cufflink," "locket," "wallet"). Each stimulus was repeated 50 times. Participants were told to respond by pressing a button as soon as they saw the stimulus appear. Following a random interstimulus interval lasting 1,400 to 1,700 ms, a target/nontarget stimulus (a string of numbers, either "11111," "22222," "33333," "44444," or "55555") was presented for 300 ms. Participants were asked to press a button if the target, "11111," appeared and to press another button if any other number string (nontarget) appeared. The target and nontargets occurred at an equal probability following probe and irrelevant stimuli. The next trial began 2,400 ms following the offset of the target/nontarget. We assumed that for guilty participants, the probe would elicit a larger P300 amplitude than an irrelevant stimulus because participants should recognize this crime-relevant item. For the innocent group, P300 amplitudes in response to the probe should be indistinguishable from those for irrelevant stimuli because the innocent participants never experienced the lab-based crime. For the two guilty groups, larger P300s to the probe than to irrelevant stimuli would suggest that the participant was knowledgeable of the crime.

After the CIT session, all participants finished a seven-block aIAT (for details, see the Supplemental Material). The critical blocks were Blocks 3 and 4 and Blocks 6 and 7. During Blocks 3 and 4, participants pressed "E" on a standard keyboard for either logically true sentences (e.g., "I am in front of a computer") or ring-relevant sentences (e.g., "I took a ring from the professor's office"; Ring+True key assignment); they pressed "I" for either logically false sentences (e.g., "I am playing football") or name-relevant sentences (e.g., "I signed my name on a poster board"; Name+False key assignment). For guilty participants, Blocks 3 and 4 were congruent (i.e., all sentences classified with the same key were true, and all classified with the other key were false), but for innocent participants, Blocks 3 and 4 were incongruent (the same keys were used to classify both true and false sentences). During Blocks 6 and 7, participants pressed "E" for either true or name-relevant sentences and "I" for either false or ring-relevant sentences (Ring+False/Name+True key assignment). These blocks were incongruent for guilty participants but congruent for innocent participants. The order of the blocks was always as described, as retaining a fixed order facilitates exploratory ERP-aIAT correlation analyses (Hu & Rosenfeld, 2012).

After the experiment, we asked participants in the two guilty conditions to rate their nervousness during the crime and their motivation to beat the CIT, and also whether they tried to distort the results of the aIAT. Participants in the suppressed-guilt group rated their compliance with the suppression instructions (e.g., how frequently they intentionally recalled the crime during the CIT; see the Supplemental Material).

EEG data acquisition

Continuous EEGs were recorded using Ag/AgCl electrodes attached to the Fz, Cz, and Pz sites according to the international 10-20 system. Scalp electrodes were referenced to linked mastoids. Electrode impedance was kept below 5 kΩ. Electrooculogram (EOG) data were recorded differentially via Ag/AgCl electrodes placed diagonally above and below the right eye to record vertical and horizontal eye movements as well as eyeblinks. EOG and EEG voltages were classified as artifacts if they exceeded 75 μV, and data from trials with artifacts were rejected. The forehead was connected to the chassis of the isolated side of the amplifier system (the ground). Signals were passed through Grass P511K amplifiers (Warwick, RI) with a 30-Hz low-pass filter and a 0.3-Hz high-pass filter (3 db). Amplifier output was passed through a 16-bit analog-to-digital converter with a sampling rate of 500 Hz.

All time windows and locations for measuring ERPs were chosen a priori based on previous ERP literature in memory detection and suppression (Bergström et al., 2009; Hu, Pornpattananangkul, & Rosenfeld, 2013; Soskins, Rosenfeld, & Niendam, 2001). We examined three ERPs: N200, P300, and late posterior negativity (LPN). The N200 was measured at the Fz electrode site and the P300 and LPN were measured at the Pz site based on their typical scalp distributions (Bergström et al., 2009; Hu et al., 2013; Soskins et al., 2001). All ERP amplitudes were measured relative to a prestimulus 100-ms baseline. The N200 was calculated by determining the mean of the most-negative 100-ms segment during the 200- to 400-ms poststimulus time window. The P300 was calculated by determining the mean of the most-positive 100-ms segment during the 300- to 800-ms poststimulus time window. This is also referred to as the base-peak P300. The LPN was calculated by determining the mean of the most-negative 100-ms segment from the P300 latency to 1,500 ms, the end of the ERP epoch. We further subtracted the LPN from the P300 to calculate a combined peak-to-peak measure.

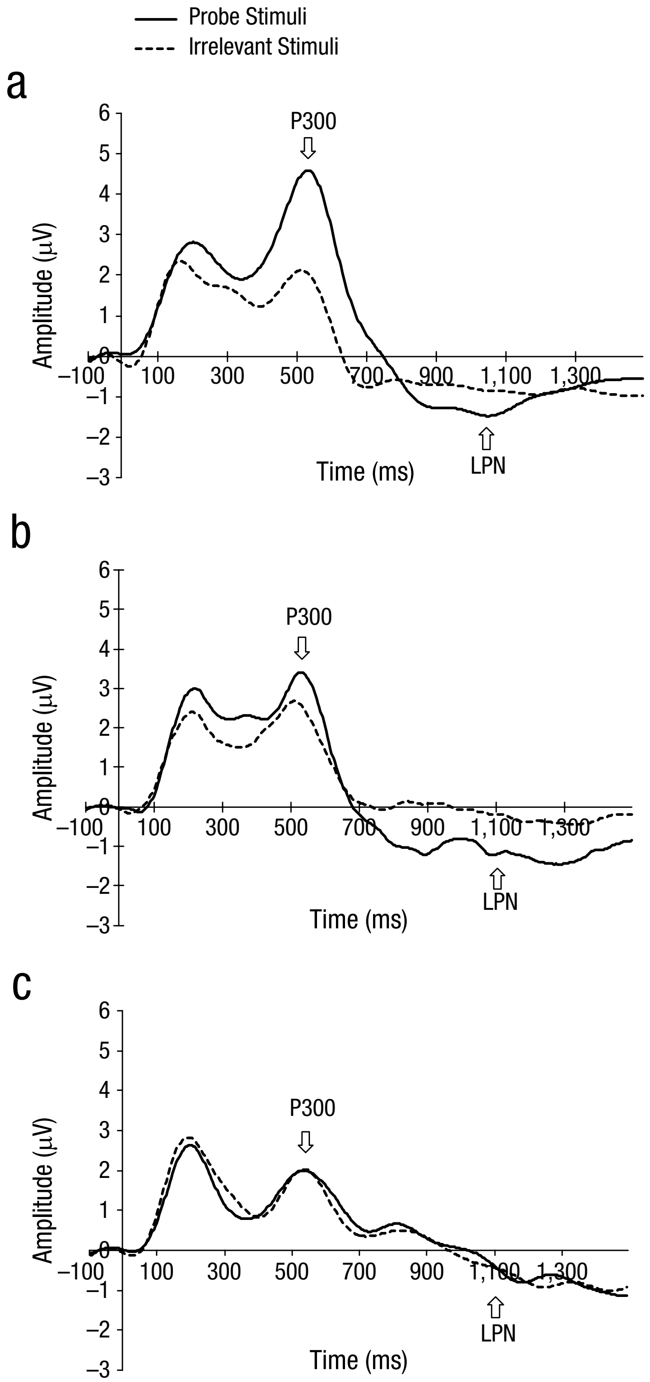

Fig. 1. Grand average event-related potentials for the (a) standard-guilt, (b) suppressed-guilt, and (c) innocent groups in response to each stimulus type. Data were recorded at the Pz site. Two components of interest—the P300 and late posterior negativity (LPN)—are indicated.

We conducted additional analyses with different ERP time windows and quantification methods to establish the replicability of the current findings; results remained the same as those reported here (see the Supplemental Material).

Data analyses

A *D* score was calculated for responses on the aIAT (Agosta & Sartori, 2013; Greenwald, Nosek, & Banaji, 2003; for details, see the Supplemental Material). A positive *D* score suggests that participants tended to associate crime-relevant sentences with truth (implying that they were guilty), whereas a negative *D* score suggests that participants tended to associate innocent sentences with truth (implying that they were innocent).

We conducted receiver-operating-characteristic (ROC) analyses to estimate the extent to which guilty participants are discriminable from innocent participants based on ERP data from the CIT. The area under the curve (AUC) is a threshold-independent indicator of the discrimination efficiency of a test that considers both sensitivity (i.e., hits) and specificity (i.e., correct rejections). The AUC represents the degree of separation between the distributions of the dependent measures in the two guilty groups and the innocent group. It varies between 0 and 1, with .5 indicating chance performance and 1 indicating perfect classification.

Results

For within-subjects analyses of variance (ANOVAs), we report 95% confidence intervals (CIs; 1.96 × standard error of the mean) following the procedures of Loftus and Masson (1994). For all ERP analyses, we conducted 3 (group: standard-guilt vs. suppressed-guilt vs. innocent; between subjects) × 2 (stimulus type: probe vs. irrelevant; within subjects) mixed ANOVAs.

In analyses of the N200, neither group, stimulus type, nor their interaction were significant, all Fs < 1.00, ps > .30, η_p^2s < .03. For the P300, there was a significant main effect of stimulus type, $F(1, 75) = 15.16$, $p < .001$, $\eta_p^2 = .168$. Probe stimuli elicited significant larger P300 amplitudes (M = 3.99 µV, 95% CI = [3.80, 4.18]) than did irrelevant stimuli (M = 3.30 µV, 95% CI = [3.11, 3.50]). Critically, the interaction between group and stimulus type was significant, $F(2, 75) = 9.95$, $p < .001$, $\eta_p^2 = .21$ (see Figs. 1 and 2).

Planned paired-samples *t* tests comparing probe and irrelevant stimuli showed that among participants in the standard-guilt group, probe stimuli elicited a significantly larger P300 amplitude (M = 4.99 µV, 95% CI = [4.66, 5.32]) than did irrelevant stimuli (M = 3.23 µV, 95% CI = [2.90, 3.57]), $t(25) = 5.19$, $p < .001$. Among participants in the suppressed-guilt group, however, no significant P300 differences between probe stimuli (M = 3.94 µV, 95% CI = [3.60, 4.28]) and irrelevant stimuli (M = 3.56 µV, 95% CI = [3.22, 3.90]) were found, $t(25) = 1.11$, $p = .280$. Comparing the P300 difference between the two types of stimuli (probe – irrelevant) in the standard-guilt and suppressed-guilt groups revealed a large effect of suppression

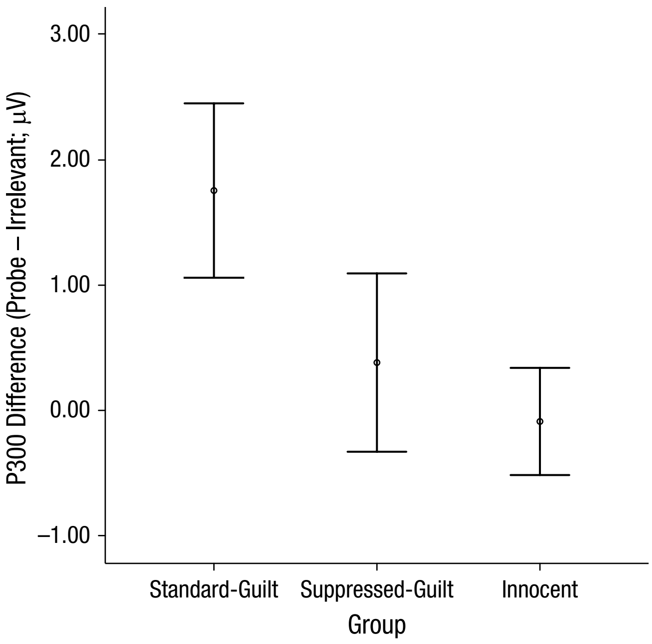

Fig. 2. Mean difference in P300 amplitude in response to probe and irrelevant stimuli, separately for the three groups. Error bars indicate 95% confidence intervals.

(Cohen's d = 0.79). Among innocent participants, there was no significant P300 difference between probe stimuli (M = 3.03 μV, 95% CI = [2.82, 3.23]) and irrelevant stimuli (M = 3.12 μV, 95% CI = [2.91, 3.32]), $t(25) = −0.43$, p = .674. Moreover, the group-by-stimulus-type interaction was not significant, which confirms that participants in the suppressed-guilt group could not be distinguished from those in the innocent group, $F(1, 50) = 1.36$, p = .249, η_p^2 = .026. The main effect of group was not significant, $F(2, 75) = 2.33$, p = .104, η_p^2 = .06.

The comparable P300 amplitudes to probe and irrelevant stimuli among participants in the suppressed-guilt group confirmed our hypothesis that suppression reduced retrieval-relevant P300 amplitudes to probes. Because this null result was central to our hypothesis, we employed Bayesian analyses to calculate the probability that, given the observed data, the null hypothesis (H_0) was true (i.e., there would be no P300 differences between probe and irrelevant stimuli among participants in the suppressed-guilt group). This probability was indexed by the formula $p(H_0 | D)$. Following the procedure recommended by Rouder, Speckman, Sun, Morey, and Iverson (2009), we showed that given our t value (1.11) and sample size (26), the odds ratio that favors the null hypothesis over the alternative hypothesis is 3.71; $p(H_0 | D) = .79$.

For the LPN, we found a main effect of stimulus type, $F(1, 75) = 33.39$, $p < .001$, η_p^2 = .308. Probe stimuli elicited a larger (i.e., more negative) LPN (M = −2.43 μV, 95% CI = [−2.60, −2.27]) than did irrelevant stimuli (M = −1.51 μV, 95% CI = [−1.68, −1.35]). The stimulus-type-by-group interaction was significant, $F(2, 75) = 5.31$, p = .007, η_p^2 = .124: Probe stimuli (M = −2.60 μV, 95% CI = [−2.93, −2.26]) elicited larger LPN amplitudes than did irrelevant stimuli (M = −1.76 μV, 95% CI = [−2.09, −1.42]) among standard-guilt participants, $t(25) = −2.47$, p = .021; the same pattern was found among suppressed-guilt participants (probe: M = −2.60 μV, 95% CI = [−2.85, −2.35]; irrelevant: M = −1.01 μV, 95% CI = [−1.26, −0.76]), $t(25) = −6.23$, $p < .001$. However, there were no LPN differences between probe stimuli (M = −2.10 μV, 95% CI = [−2.32, −1.89]) and irrelevant stimuli (M = −1.78 μV, 95% CI = [−1.99, −1.56]) among innocent participants, $t(25) = −1.52$, p = .142. No effect of group was found, $F(2, 75) = 0.35$, p = .703, η_p^2 = .009.

Next, we examined the P300 – LPN combined measure, and we found a significant main effect of stimulus type, $F(1, 75) = 43.20$, $p < .001$, η_p^2 = .365. Probe stimuli elicited a larger P300 – LPN (M = 6.42 μV, 95% CI = [6.16, 6.68]) than did irrelevant stimuli (M = 4.82 μV, 95% CI = [4.56, 5.08]). A significant group-by-stimulus-type interaction was also found, $F(2, 75) = 8.36$, p = .001, η_p^2 = .182. Probe stimuli elicited larger P300 – LPN measures than irrelevant stimuli among participants in both the standard-guilt group (probe: M = 7.59 μV, 95% CI = [7.12, 8.05]; irrelevant: M = 4.99 μV, 95% CI = [4.53, 5.46]), $t(25) = 5.48$, $p < .001$, and the suppressed-guilt group (probe: M = 6.54 μV, 95% CI = [6.07, 7.02]; irrelevant: M = 4.57 μV, 95% CI = [4.09, 5.05]), $t(25) = 4.06$, $p < .001$. Among innocent participants, no significant difference was found between amplitudes in response to probe stimuli (M = 5.13 μV, 95% CI = [4.86, 5.40]) and irrelevant stimuli (M = 4.89 μV, 95% CI = [4.62, 5.16]), $t(25) = 0.88$, $p > .30$. There was no main effect of group, $F(2, 75) = 1.55$, p = .219, η_p^2 = .04. The ERP-based individual diagnoses (i.e., the ROC analysis) showed a highly similar pattern as the foregoing group analyses (see the Appendix).

Turning to the aIAT, we first excluded 1 participant from the suppressed-guilt group because he indicated on his postexperiment questionnaire that he intentionally suppressed crime memories during the aIAT, which left 25 participants (results remained the same regardless of this exclusion). Moreover, because EEG artifacts would not affect participants' aIAT performance, an additional analysis was conducted regardless of whether participants had EEG artifacts (n = 34 in the standard-guilt group, n = 29 in the suppressed-guilt group). Results were the same in these two analyses. We report the first analysis here, as it allows for exploratory correlation analyses between ERP and aIAT data.

A one-way ANOVA showed that D scores in the three groups were significantly different from each other, $F(2, 74) = 27.19$, $p < .001$. Because innocent participants wrote their initials without enacting the lab-based crime, their

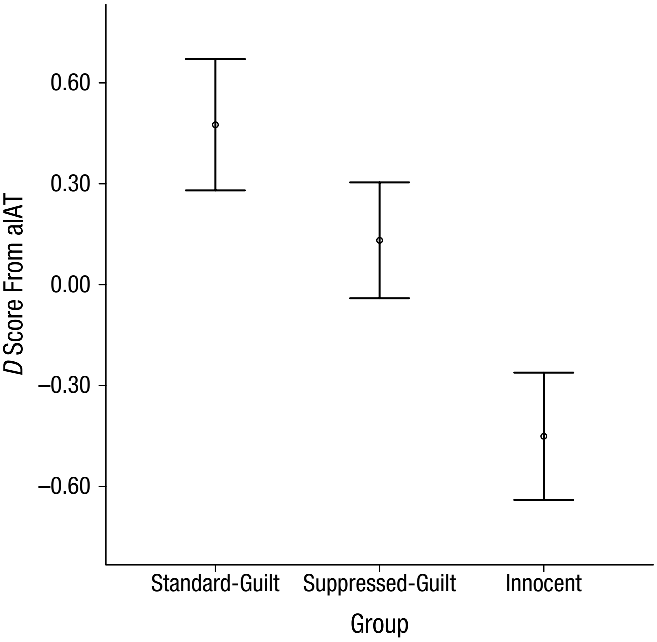

Fig. 3. Mean *D* score from the autobiographical Implicit Association Test (aIAT) for the three groups. Error bars indicate 95% confidence intervals. *D* scores above zero suggest that the crime-related memories are true; *D* scores below zero suggest that the non-crime-related memories are true.

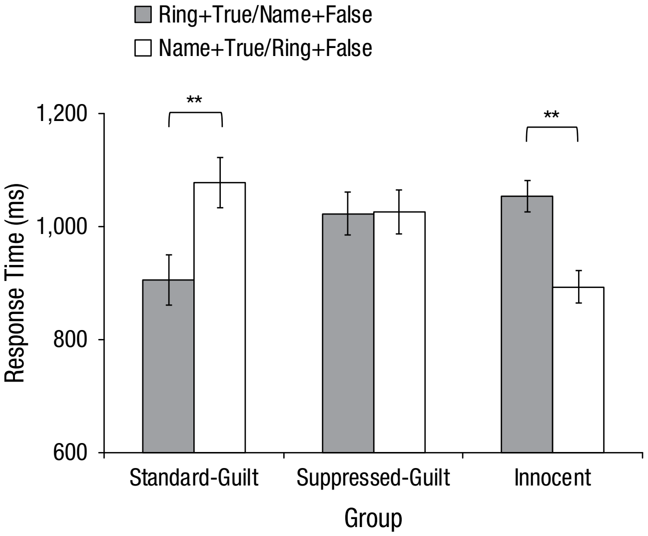

Fig. 4. Mean response time from the Ring+True/Name+False and Name+True/Ring+False blocks in the autobiographical Implicit Association Test. Error bars indicate 95% confidence intervals. The Ring+True block was a congruent block for guilty participants yet an incongruent block for innocent participants, whereas the Name+True block was a congruent block for innocent participants but an incongruent block for guilty participants. Asterisks indicate a significant difference between groups ($^{**}p < .001$).

D scores were negative ($M = -0.45$, 95% CI = [−0.63, −0.27]). Most important, *D* scores for suppressed-guilt participants ($M = 0.13$, 95% CI = [−0.03, 0.29]) were significantly smaller than for standard-guilt participants ($M = 0.47$, 95% CI = [0.29, 0.66]), $t(50) = 2.71$, $p = .009$, Cohen's $d = 0.76$, despite both groups having experienced the lab-based crime (Fig. 3).

Because participants in the standard-guilt group finished a CIT before the aIAT, the CIT may have reminded them of the crime and therefore artificially increased the aIAT effect. To address this concern, we compared the aIAT from the suppressed-guilt group with similar aIATs that were not preceded by CITs (baseline aIATs in Hu, Rosenfeld, & Bodenhausen, 2012; first aIAT administrations summarized in Agosta & Sartori, 2013). Using these aIATs as a baseline, results still showed that suppressing memories led to significantly reduced *D* scores (the nonoverlapping 95% CIs indicate significant differences)—present experiment: $M = 0.13$, 95% CI = [−0.03, 0.29]; Hu et al. (2012, $N = 64$): $M = 0.49$, 95% CI = [0.40, 0.58]; Agosta and Sartori (2013; $N = 412$): $M = 0.58$, 95% CI = [0.41, 0.73]. Thus, the effect of suppression on the aIAT is unlikely to be attributable to artificially increased aIAT scores when participants first completed the CIT.

To better understand the reduction of *D* scores and exclude concerns that participants distorted the aIAT results by intentionally slowing their responses, we analyzed response times from the aIAT. A 3 (group: standard-guilt vs. suppressed-guilt vs. innocent; between subjects) × 2 (block: congruent vs. incongruent; within subjects) mixed ANOVA showed that the group-by-block interaction was significant, $F(2, 74) = 19.04$, $p < .001$, $\eta_p^2 = .34$. Follow-up analyses showed that among innocent participants, the Name+True ($M = 893.67$ ms, 95% CI = [865.54, 921.80]) versus Ring+True ($M = 1{,}053.99$ ms, 95% CI = [1,025.86, 1,082.12]) congruence effect was significant, $t(25) = -5.59$, $p < .001$. Among standard-guilt participants, the Ring+True ($M = 905.63$ ms, 95% CI = [860.96, 950.30]) versus Name+True ($M = 1{,}077.73$ ms, 95% CI = [1,033.06, 1,122.40]) congruence effect was also significant, $t(25) = -3.78$, $p = .001$. In contrast, among suppressed-guilt participants, there was no Ring+True ($M = 1{,}023.18$ ms, 95% CI = [985.11, 1,061.25]) versus Name+True ($M = 1{,}026.09$ ms, 95% CI = [988.03, 1,064.16]) congruence effect, $t(24) = -0.08$, $p > .90$ (see Fig. 4). Employing the same Bayesian analysis procedure as for the P300, we found that the odds ratio favoring this null hypothesis to the alternative hypothesis was 6.48, $p(H_0 | D) = .87$.[1]

Discussion

People can consciously suppress unwanted memories, but a century-old question is whether such suppressed memories can nevertheless influence people's behavior in a less conscious, more automatic manner than they would if they had not been suppressed. We provide novel evidence that not only can people suppress neural activity underlying retrieval of sensorimotor-rich memories, but also this suppression limits subsequent automatic influences of these memories.

The amplitude of the P300 has been linked to conscious recollection of episodic memories, especially the richness of such recollection (Paller et al., 1995; Rugg & Curran, 2007; Vilberg et al., 2006). An attenuated P300 response to crime-relevant details provides direct neural evidence that people can voluntarily terminate the retrieval of unwanted sensorimotor-rich memories. Critically, our standard-guilt group did not receive intentional retrieval instructions. Thus, the attenuated P300 in the suppressed-guilt group was the result of down-regulation of retrieval-related neural activity rather than up-regulation seen in the standard-guilt group, which supports the notion that inhibitory processes can directly override automatic retrieval (Anderson & Hanslmayr, 2014).

Despite their success at terminating recollection, participants in the suppressed-guilt group nevertheless revealed their guilt via enlarged LPNs. This LPN is dissociable from the recollection-sensitive P300s (Rugg, Schloerscheidt, Doyle, Cox, & Patching, 1996) and may indicate response-monitoring processes (Johansson & Mecklinger, 2003). Here, participants in the suppressed-guilt group voluntarily suppressed the criminal memories associated with the crime-relevant details, which would otherwise trigger automatic retrieval. The enlarged LPN may reflect the enhanced need to monitor response conflict between top-down suppression and automatic recognition processes.

Another possible suppression-sensitive neural signal is the frontal N200, which indicates top-down inhibition and predicts later forgetting (Bergström et al., 2009). However, this N200 was absent here. Because suppressed-guilt participants engaged in suppression throughout the whole memory test, such continuous suppression may be difficult to detect in a trial-specific manner (Bergström et al., 2013). In contrast, when intentional retrieval and suppression trials were intermixed on a trial-by-trial basis that also involved task switching, this suppression-sensitive N200 was more evident (Bergström et al., 2009).

Unwanted memories can intrude into consciousness automatically despite goal-directed suppression. Such intrusions can be purged from consciousness by retrieval suppression, which eventually weakens memory representations (Levy & Anderson, 2012). Moreover, suppressing memories of visual scenes can make them less identifiable in perceptual priming tasks (Gagnepain et al., 2014; Kim & Yi, 2013). Here, we obtained similar findings: Top-down suppression limited the automatic influence of previously suppressed memories, even when to-be-suppressed memories were sensorimotor-rich and self-referential (Cabeza & St Jacques, 2007). Indeed, during the aIAT, participants in the suppressed-guilt group behaved as if they had not experienced the lab-based crime. Together with previous research, this finding suggests that retrieval suppression can render unwanted memories both less consciously accessible and less likely to exert automatic, implicit influences on behavior.

The finding that criminal suspects can willfully terminate retrieval of criminal memories and their associated brain activity is problematic for neuroscience-based memory assessments. Nevertheless, suppression may leave its neural traces (in the LPN), which suggests that criminals employing this countermeasure may still be identifiable using some memory-detection protocols. Future research should test whether guilty individuals who engage in a memory-suppression strategy can be detected via functional MRI, because suppression attempts engage the dorsolateral prefrontal cortex (Anderson et al., 2004). It is also important to assess whether suppression can reduce automatic influences of arousing, traumatic autobiographical memories. Tackling these intriguing questions has implications for treatment of psychopathologies that are characterized by automatic intrusion of unwanted memories.

Appendix

Examination of the base-peak P300 allowed us to successfully differentiate standard-guilt from innocent participants (AUCs = .84, $p < .001$), as well as from suppressed-guilt participants (AUC = .74, $p = .003$). However, we could not differentiate between suppressed-guilt and innocent participants using the P300 (AUC = .57, $p = .37$). Thus, suppression renders it ineffective to identify guilty participants using the P300. However, the LPN among the suppressed-guilt group still distinguished them from innocent participants (AUC = .76, $p = .001$). Combining the P300 and LPN in a peak-to-peak manner (i.e., P300 – LPN; Soskins et al., 2001) can enable one to discriminate guilty and innocent populations regardless of whether or not they suppressed memories (AUCs > .70, ps < .01; see Table A1).

Analysis of the postexperiment questionnaires revealed no differences between motivations to beat the test or nervousness during the lab-crime ratings between the two guilty groups (ps > .12). Ratings of participants in the standard-guilt group suggested that the crime memories came to mind relatively automatically ($M = 3.62$, $SD = 0.28$, on a scale from 0 to 6; see the Supplemental Material), but less automatically than in previous research (Bergström et al., 2013, obtained a mean of 3.90, $SD = 0.06$, on a scale from 1 to 4). This discrepancy can be

Table A1. Areas Under the Curve From the Receiver-Operating-Characteristic Analyses

Group comparison	P300	LPN	P300 – LPN
Standard-guilt vs. innocent	.84 [.72, .96]*	.60 [.45, .76]	.80 [.69, .92]*
Suppressed-guilt vs. innocent	.57 [.42, .73]	.76 [.63, .89]*	.73 [.59, .87]*
Standard-guilt vs. suppressed-guilt	.74 [.60, .88]*	.63 [.48, .78]	.56 [.40, .72]

Note: Values in brackets are 95% confidence intervals.
*$p < .01$.

ascribed to different lab-based-crime procedures. In Bergström et al. (2013), participants encoded memories during a computer-based crime-simulation task, wherein they navigated a virtual environment and vividly imagined committing a burglary. This simulation task was designed to lead to rich and elaborate memories. Here, in contrast, we adopted an incidental encoding scenario that is much more relevant to real-life crime-memory detection but that may discourage in-depth encoding or rehearsal of crime details because of time pressure. The real-life and simulation-based procedures could yield different levels of encoding depth of to-be-suppressed memories, which could account for differences in both suppression ERP effects and automaticity ratings between the two studies.

Finally, in addition to the hypothesis-driven analyses described in the main text, exploratory analyses indicated that (a) suppression may have affected automatic aspects of aIAT performance more than controlled aspects, (b) suppressed-guilt participants' aIAT performance could not be predicted by any of the measured ERP components, and (c) the P300 and LPN components were indeed orthogonal (for details, see the Supplemental Material).

Author Contributions

X. Hu developed the study concept. All authors contributed to the experimental design. X. Hu collected and analyzed the data and drafted the first version of the manuscript. Z. M. Bergström, G. V. Bodenhausen, and J. P. Rosenfeld provided critical revisions. All authors approved the final version of the manuscript.

Acknowledgments

We thank Mike Anderson for providing the screening questionnaires.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Part of this research was supported by an American Psychological Association Dissertation Research Award to X. Hu.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data and materials have been made publicly available via Open Science Framework and can be accessed at https://osf.io/ptmqx/. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/view/ and http://pss.sagepub.com/content/25/1/3.full.

Note

1. See the Appendix for results from the postexperimental questionnaire and exploratory analyses.

References

Agosta, S., & Sartori, G. (2013). The autobiographical IAT: A review. *Frontiers in Psychology*, *4*, Article 519. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00519/full

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 366–369. doi:10.1038/35066572

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*, 279–292. doi:10.1016/j.tics.2014.03.002

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., . . . Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*, 232–235. doi:10.1126/science.1089504

Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*, *76*, 450–460. doi:10.1016/j.neuron.2012.07.025

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive top–down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *Journal of Cognitive Neuroscience*, *27*, 96–111. doi:10.1162/jocn_a_00696

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, *94*, 1–11. doi:10.1016/j.biopsycho.2013.04.012

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*, 726–737. doi:10.1016/j.neuroimage.2009.06.051

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*, 219–227. doi:10.1016/j.tics.2007.02.005

Depue, B. E. (2012). A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. *Neuroscience & Biobehavioral Reviews*, *36*, 1382–1399. doi:10.1016/j.neubiorev.2012.02.012

Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and neural oscillations during volitional suppression of memory retrieval. *Journal of Cognitive Neuroscience*, *25*, 1624–1633. doi:10.1162/jocn_a_00418

Farah, M. J., Hutchinson, J. B., Phelps, E. A., & Wagner, A. D. (2014). Functional MRI-based lie detection: Scientific and societal challenges. *Nature Reviews Neuroscience*, *15*, 123–131. doi:10.1038/Nrn3665

Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences, USA*, *111*, E1310–E1319. doi:10.1073/pnas.1311468111

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216. http://dx.doi.org/10.1037/0022-3514.85.2.197

Hu, X., Pornpattananangkul, N., & Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology*, *50*, 454–464. doi:10.1111/psyp.12018

Hu, X., & Rosenfeld, J. P. (2012). Combining the P300-complex trial-based concealed information test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. *Psychophysiology*, *49*, 1090–1100. doi:10.1111/j.1469-8986.2012.01389.x

Hu, X., Rosenfeld, J. P., & Bodenhausen, G. V. (2012). Combating automatic autobiographical associations: The effect of instruction and training in strategically concealing information in the autobiographical Implicit Association Test. *Psychological Science*, *23*, 1079–1085. doi:10.1177/0956797612443834

Johansson, M., & Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: Action monitoring and retrieval of attribute conjunctions. *Biological Psychology*, *64*, 91–117. doi:10.1016/s0301-0511(03)00104-2

Kim, K., & Yi, D. J. (2013). Out of mind, out of sight: Perceptual consequences of memory suppression. *Psychological Science*, *24*, 569–574. doi:10.1177/0956797612457577

Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *The Journal of Neuroscience*, *32*, 16785–16794. doi:10.1523/JNEUROSCI.2640-12.2012

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490. doi:10.3758/BF03210951

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*, 879–904. doi:10.1111/psyp.12239

Noreen, S., & MacLeod, M. D. (2013). It's all in the detail: Intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 375–393. doi:10.1037/a0028888

Paller, K. A., Kutas, M., & McIsaac, H. K. (1995). Monitoring conscious recollection via the electrical activity of the brain. *Psychological Science*, *6*, 107–111. doi:10.1111/j.1467-9280.1995.tb00315.x

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*, 118–134. doi:10.1016/j.ijpsycho.2013.08.012

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*, 251–257. doi:10.1016/j.tics.2007.04.004

Rugg, M. D., Schloerscheidt, A. M., Doyle, M. C., Cox, C. J., & Patching, G. R. (1996). Event-related potentials and the recollection of associative information. *Cognitive Brain Research*, *4*, 297–304. doi:10.1016/S0926-6410(96)00067-5

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501–518. doi:10.1037//0278-7393.13.3.501

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: Complex vs. simple paradigms. *International Journal of Psychophysiology*, *40*, 173–180. doi:10.1016/s0167-8760(00)00154-9

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, *1122*, 161–170. doi:10.1016/j.brainres.2006.09.023