Running head: SUPPRESSING UNWANTED AUTOBIOGRAPHICAL MEMORIES

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

Xiaoqing Hu[1,2 CA], Zara M. Bergström[3], Galen V. Bodenhausen[1], J. Peter Rosenfeld[1]

1, Department of Psychology, Northwestern University, Evanston, US

2, Department of Psychology, University of Texas at Austin, Austin, US

3, School of Psychology, University of Kent, UK

Corresponding Author:

Xiaoqing Hu,

Department of Psychology, University of Texas, Austin, Austin, TX, 78712

E-mail: xqhu@utexas.edu

Word Count: Abstract: 149 words

Introduction + Discussion: 450+538 = 988 words.

Method + Results + Figures + Table: 2992 words

References: 30

Abstract

The present study investigated the extent to which people can suppress unwanted autobiographical memories in a mock crime memory detection context. Participants encoded sensorimotor-rich memories by enacting a lab crime (stealing a ring) and received direct suppression instructions so as to evade guilt detection in a brainwave-based concealed information test. Aftereffects of suppression on automatic memory processes were measured in an autobiographical implicit association test (aIAT). Results showed that suppression attenuated brainwave activity (P300) that is associated with crime-relevant memory retrieval, rendering innocent and guilty/suppression participants indistinguishable. However, guilty/suppression and innocent participants 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. Lastly, extending recent findings that suppression can impair implicit memory processes; we provide novel evidence that suppression reduces automatic cognitive biases that are otherwise associated with actual autobiographical memories.

Key Words: memory suppression, memory detection, autobiographical memory, P300, autobiographical implicit association test, neuroscience and law.

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

The automatic intrusion of unwelcome memories can sting. People commonly rely on inhibitory control to prevent unwanted memories from intruding, which reduces explicit recall of such memories (Anderson & Green, 2001). Neuroimaging research suggests that suppressing previously encoded words/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 can illuminate theoretical issues in cognitive control as well as offer practical implications in translational fields such as neurolaw regarding neuroscientific approaches to guilt detection (Farah, Hutchinson, Phelps & Wagner, 2014).

We investigated these issues in a memory detection context. Participants were asked to suppress sensorimotor-rich memories that were encoded during a lab crime. We hypothesized that suppressing autobiographical memory can attenuate the P300, an event-related brain potential (ERP) indicating conscious recollection (Paller, Kutas & McIsaac, 1995; Rugg & Curran, 2007; Vilberg, Moosavi & Rugg, 2006) that has been long used in memory detection (Rosenfeld et al., 2013). Indeed, retrieval suppression can reduce P300s to previously learned

words (Bergström, deFockert & Richardson-Klavehn, 2009; Depue et al., 2013), and pictures in memory detection tests (Bergström et al., 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 potentially describe a past autobiographical activity (e.g., I took a ring) and must classify these statements in terms of their general topic (as a "ring-related" event or not). On intermixed trials, they are asked to confirm or deny unequivocally true (e.g., "I am sitting in front of a computer") or false statement (e.g., "I am climbing a mountain"). The veracity of the autobiographical statements can be inferred from the speed/accuracy of making these simple classifications (Agosta & Sartori, 2013).

Importantly, 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 et al., 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 predetermined our sample size to be 26 participants/group. This sample size was chosen because a power analysis indicated 26 participants/group were required to detect a large suppression effect (Cohen's *d*=0.8) with power =0.8 at alpha=0.05; we expected a large effect in suppressing incidentally encoded crime-relevant memories given 1) a recent meta-analysis in memory detection suggests the P300 is extremely sensitive to variations of recognition (Meijer et al., 2014), and 2) the most relevant prior memory suppression research typically produced medium to large suppression effects (Bergström et al., 2013; Gagnepain et al., 2014, Kim & Yi, 2013).This sample size is also consistent with relevant prior memory suppression studies (which typically involved 24 participants per experiment/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 for EEG artifacts (*N*=15) or not following instructions (*N*=9), see SOM). Participants were compensated with either course credit or money. Participants were additionally promised a $10 reward if an innocent outcome is obtained from the brainwave-based test. They were later given this $10 regardless of their performance. The study was approved by the Northwestern Institutional Review Board.

Procedure

Participants were randomly assigned to one of three groups (*N*=26 per group):1) a standard guilty group without any suppression instructions; 2) a suppression/guilty group given memory

suppression instructions; and 3) an innocent group without any lab crime and without suppression instructions. Except as noted, all participants completed the following: 1) they enacted either a lab crime (described below) or an innocent act (~ 10 mins); 2) an ERP-based concealed information test (CIT, ~30 mins); 3) an aIAT (~10 mins) and 4) post-experiment questionnaires for all guilty participants (~3 mins).

Lab Crime/Innocent Act: Participants in both guilty groups were instructed to enact a lab crime: to find and steal something (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 sign their name initials on a poster board near the office. They were thus unaware of any lab crime.

Memory Suppression Manipulation: Before the ERP-based CIT, participants in the suppression group received direct suppression instructions (Benoit & Anderson, 2012; Bergström et al., 2009): they should never allow the lab crime memory come to mind at all during the test, and they should not engage in distracting thoughts (see SOM).

ERP-based CIT: The present study employed the complex trial version of the CIT (see SOM), which is more countermeasure-resistant 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 (e.g. 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 this stimulus. Following a random inter-stimulus

interval lasting 1400-1700 ms, a target/non-target stimulus (a string of numbers, either 11111, 22222, 33333, 44444 or 55555) was presented for 300 ms. Participants were asked to press a button for the target "11111", and to press another button for any other number string (non-targets). The target and non-target occurred at an equal probability following probe and irrelevant stimuli. The next trial began 2400 ms following the offset of the target/non-target. The CIT assumes that for guilty participants, the probe will elicit a larger P300 than an irrelevant stimulus because they should recognize this crime-relevant item. For innocents who are unaware of the crime, the probe was never encountered; no recognition is involved. When P300s to the probe are larger than P300s to irrelevant stimuli, one can infer that the participant is knowledgeable of the crime.

RT-based aIAT: After the ERP session, all participants finished a seven-block aIAT (for details, see SOM). The critical blocks are blocks 3,4 and 6,7. During blocks 3 and 4, participants pressed keyboard button "E" for either logically true (e.g., I am in front of a computer) or Ring-relevant sentences (e.g., I took a ring from the professor's office); and they pressed button "I" for either logically false (e.g., I am playing football) or Name-relevant sentences (e.g., I signed my name on a poster board). Blocks 3 and 4 were congruent for guilty but incongruent for innocent participants. During blocks 6 and 7, participants pressed button "E" for either true or Name-relevant sentences and button "I" for either false or Ring-relevant sentences. These blocks were incongruent for guilty but congruent for innocent participants. The order of the double classification blocks was always as described above, as it facilitates exploratory ERP-aIAT correlation analyses (Hu & Rosenfeld, 2012).

Post-experiment Questionnaires: We asked all guilty participants to rate their nervousness during the crime, their motivation to beat the CIT, and whether they tried to distort the aIAT. Guilty/suppression participants rated their compliance with the suppression instructions (e.g., how frequently they intentionally recalled the crime during the CIT, see SOM).

EEG Data Acquisition

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

ERP Measurements:

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 et al., 2013; Soskins et al., 2001). We examined three ERPs: N200, P300 and late posterior negativity (LPN). The N200 was measured at Fz and the P300 and LPN at Pz 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 pre-stimulus 100 ms baseline. The N200 was calculated as the mean

of the most negative 100-ms segment during the 200-400 ms post-stimulus time window. The P300 was calculated as the mean of the most positive 100-ms segment during the 300-800 ms post-stimulus time window. This is also referred to as the base-peak P300. The LPN was calculated as the mean of the most negative 100-ms segment from the P300 latency to 1500 ms, the end of the ERP epoch. We further subtracted the LPN from the P300 (P300-minus-LPN) as a combined, peak-peak measure. 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 SOM).

D-score for the aIAT:

A D_{600} score (D-score) was calculated (Greenwald, Nosek, & Banaji, 2003; Agosta & Sartori, 2013, for details, see SOM). A positive D-score suggests participants tend to associate crime-relevant sentences with truth (implying guilt) whereas a negative D-score suggests participants tend to associate innocent sentences with truth (implying innocence).

Classification Efficiency

We conducted receiver operating characteristic (ROC) analyses to estimate the extent to which guilty participants are discriminatable from innocent participants based on the ERP-CIT. The area under the curve (AUC) is a threshold-independent indicator of the discrimination efficiency of a test, considering 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 from guilty (standard/suppression groups) and innocent participants. It varies between 0 and 1, with a chance level of 0.5 and with a perfect classification level of 1.

Results

Effect Size Report

For ANOVA analyses, we report partial eta square (η_p^2); for between-group comparisons, we report Cohen's *d* as an index of effect size. 95% Confidence Intervals (CIs) are provided with means. For within-subject comparisons, we calculated the 95% CIs (1.96 * Standard Error of Means) based on Loftus and Masson (1994).

ERPs in the CIT

For all ERP analyses, we conducted 3 (between-subject, guilty/standard vs. guilty/suppression vs. innocent) by 2 (within-subject, probe vs. irrelevant, average of all irrelevant) mixed ANOVAs.

N200: Neither Group, Stimulus type, nor their interaction were significant: all Fs<1.00, ps>.30, η_p^2s<0.03.

P300: The main effect of stimulus type was significant $F(1,75)=15.16$, $p<.001$, $\eta_p^2=0.168$. Probe elicited significant larger P300s (Means and 95% CIs: 3.99 μV, [3.80, 4.18]) than irrelevants (3.30 μV, [3.11, 3.50]). Critically, the interaction between group and stimulus type was significant $F(2, 75)=9.95$, $p<.001$, $\eta_p^2=0.21$ (see Fig. 1 & 2). Planned probe vs. irrelevant paired-sample *t*-tests showed that among guilty/standard participants, the probe stimulus elicited a significantly larger P300 than irrelevant stimuli, $t(25)=5.19$, $p<.001$, probe: 4.99 μV [4.66, 5.32] vs. irrelevant: 3.23 μV [2.90, 3.57]. Among guilty/suppression participants, however, no significant P300 differences between probe and irrelevant were found, $t(25)=1.11$, $p=.280$, probe: 3.94 μV [3.60, 4.28] vs. irrelevant 3.56 μV [3.22, 3.90]. Comparing the probe-minus-irrelevant P300 differences between guilty/standard and guilty/suppression revealed a large effect size of

suppression: Cohen's *d* =0.79. Among innocent participants, there was no significant P300 difference between probe and irrelevant $t(25)=-0.43$, $p=.674$, probe: 3.03 µV [2.82, 3.23] vs. irrelevant: 3.12 µV [2.91, 3.32]. Moreover, the suppression vs. innocent by probe vs. irrelevant interaction was not significant, confirming that guilty suppressors could not be distinguished from innocents ($F(1,50)=1.36$, $p=.249$, $\eta_p^2=0.026$). The main effect of group was not significant $F(2,75)=2.33$, $p=.104$, $\eta_p^2=0.06$.

The comparable P300s to probe and irrelevant among guilty/suppression participants confirms our hypothesis that suppression reduced retrieval-relevant P300s to probes. Because this null result is central to our hypothesis, we employed Bayesian analyses to calculate the $p(H_0|D)$, i.e., given the observed data, the probability the null hypothesis is true (no probe vs. irrelevant P300 differences among suppression participants). 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 null hypothesis (H_0) to the alternative hypothesis (H_1) is 3.71; $p(H_0|D)=0.79$.

Fig. 1: Grand average ERPs recorded at Pz. Solid Line represents Probe, dashed line represents average of all irrelevant.

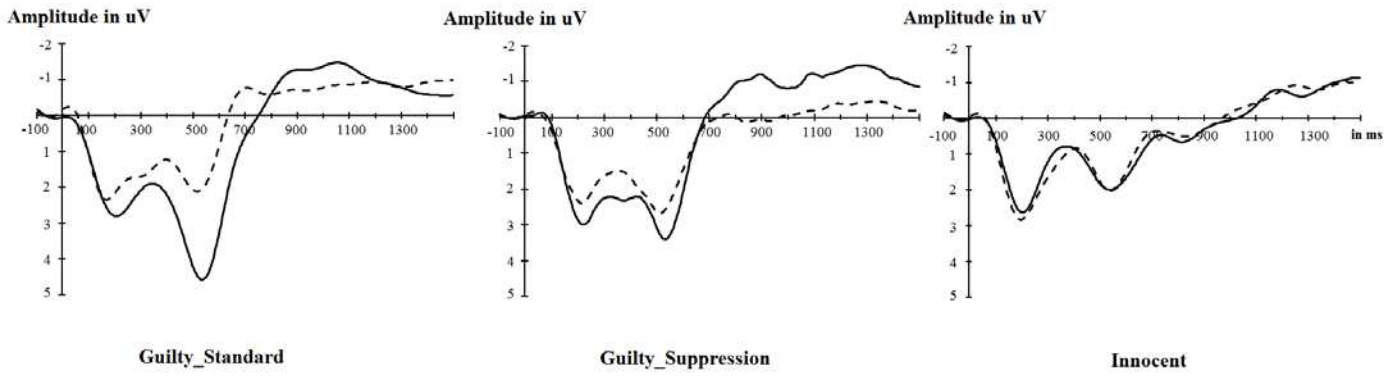
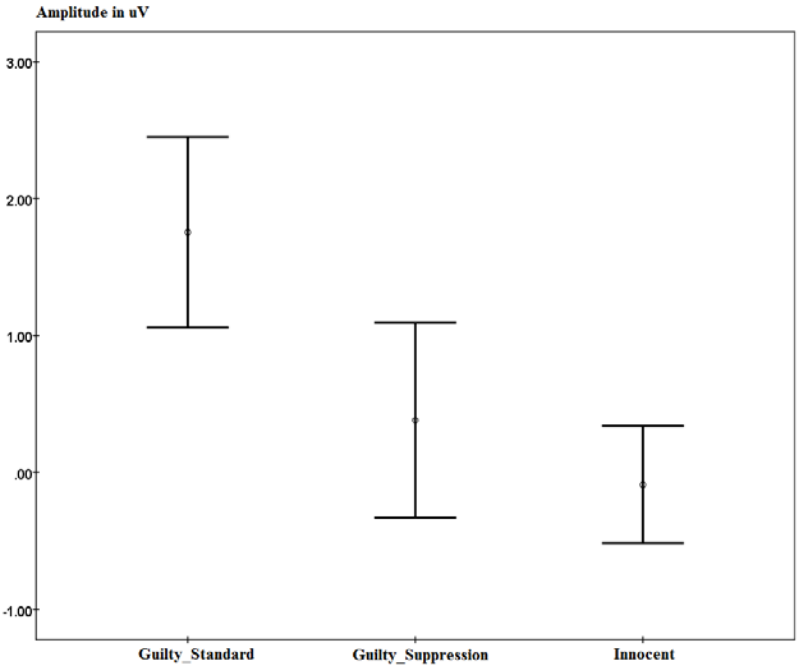


Fig. 2 Probe-minus-irrelevant P300s from all three groups. Error bars indicate 95% CIs. Zero on the Y-axis indicates the probe and irrelevant are not different from each other.



LPN: We found a main effect of stimulus type: $F(1,75)=33.39$, $p<.001$, $\eta_p^2=0.308$. The probe elicited a larger (i.e., more negative) LPN (-2.43 µV [-2.60, -2.27]) than irrelevant (-1.51

μV [-1.68, -1.35]). The stimulus by group interaction was significant $F(2,75)=5.31$, $p=.007$, $\eta_p^2=0.124$: probe elicited larger LPN than irrelevant among guilty/standard participants, ($t(25)=-2.47$, $p=.021$, probe: -2.60 μV [-2.93, -2.26] vs. irrelevant -1.76 μV [-2.09, -1.42]), and among guilty/suppression participants ($t(25)=-6.23$, $p<.001$, probe: -2.60 μV [-2.85, -2.35] vs. irrelevant -1.01 μV [-1.26. -0.76]). However, there were no probe vs. irrelevant LPN differences among innocent participants: $t(25)=-1.52$, $p=.142$, probe: -2.10 μV [-2.32, -1.89] vs. irrelevant -1.78 μV [-1.99, -1.56]. No group effect was found: $F(2,75)=0.35$, $p=.703$, $\eta_p^2=0.009$.

Combining P300 and LPN: A significant main effect of stimulus type was found, $F(1,75)=43.20$, $p<.001$, $\eta_p^2=0.37$. Probe elicited a larger P300-minus-LPN (6.42 μV [6.16, 6.68]) than irrelevant (4.82 μV [4.56, 5.08]). A significant group by stimulus interaction was also found: $F(2,75)=8.36$, $p=.001$, $\eta_p^2=0.18$. Probe elicited larger P300-minus-LPN than irrelevant among both guilty/standard (probe: 7.59 μV [7.12, 8.05] vs. irrelevant 4.99 μV [4.53, 5.46], $t(25)=5.48$, $p<.001$) and guilty/suppression participants (probe 6.54 μV [6.07, 7.02] vs. irrelevant: 4.57 μV [4.09, 5.05], $t(25)=4.06$, $p<.001$). No probe vs. irrelevant difference was found among innocent participants (probe: 5.13 μV [4.86, 5.40] vs. irrelevant 4.89 μV [4.62, 5.16], $t(25)=0.88$, $p>.30$). There was no group main effect: $F(2,75)=1.55$, $p=.219$, $\eta_p^2=0.04$.

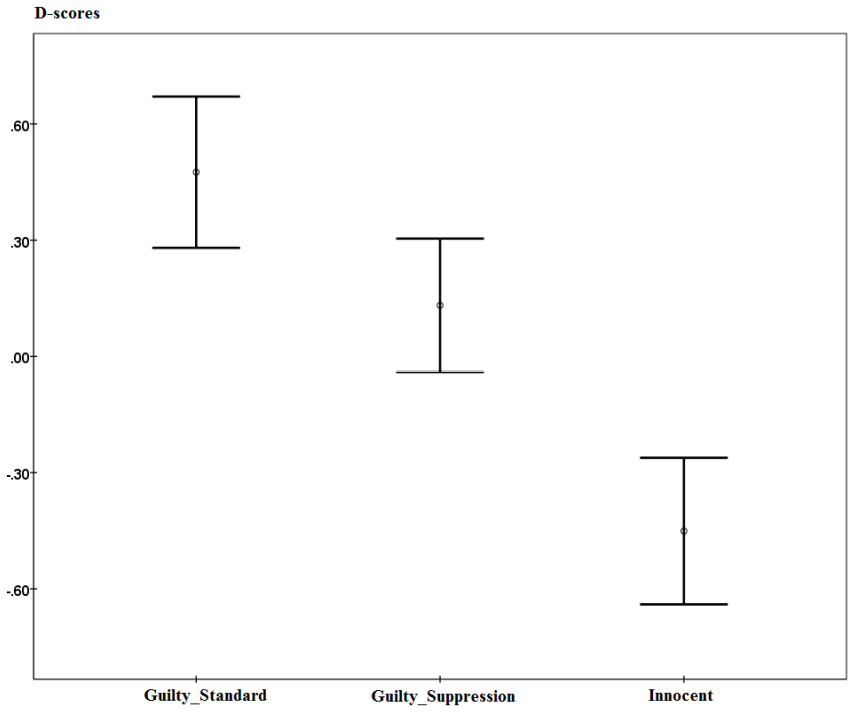
Influence of Suppression on the aIAT

One participant from the suppression group was excluded because he intentionally suppressed crime memories during the aIAT based on his post-experiment questionnaire, leaving 25 participants (results remain the same regardless of this exclusion). Moreover, because EEG artifacts will not affect participants' aIAT performance, an additional analysis was conducted with participants regardless of EEG artifacts ($N=34$ in the standard group, 29 in the suppression

group). Results were the same in these two analyses. We report the first analysis here as it allows for exploratory ERP-aIAT correlation analyses.

D-score Analyses: A one-way ANOVA on the D-scores from the three groups revealed that D-scores were significantly different from each other. $F(2, 74)= 27.19$, $p<.001$. Because innocent participants signed their name without enacting the lab crime, their D-scores were negative Mean and 95% CI: -0.45, [-0.63, -0.27]. Most importantly, D-scores for guilty/suppression participants (0.13 [-0.03, 0.29]) were significantly smaller than for guilty/standard participants (0.47, [0.29, 0.66]), $t(50)=2.71$, $p=.009$, Cohen's $d = 0.75$, despite both groups having experienced the lab crime (Fig. 3).

Fig. 3: D-scores from the aIAT for all three groups. Error bars indicate 95% CIs. D-scores above zero suggests that the crime-relevant memories (e.g., I took a ring) are true; D-scores below zero suggests that the innocent-relevant memories (e.g., I signed my name) are true.

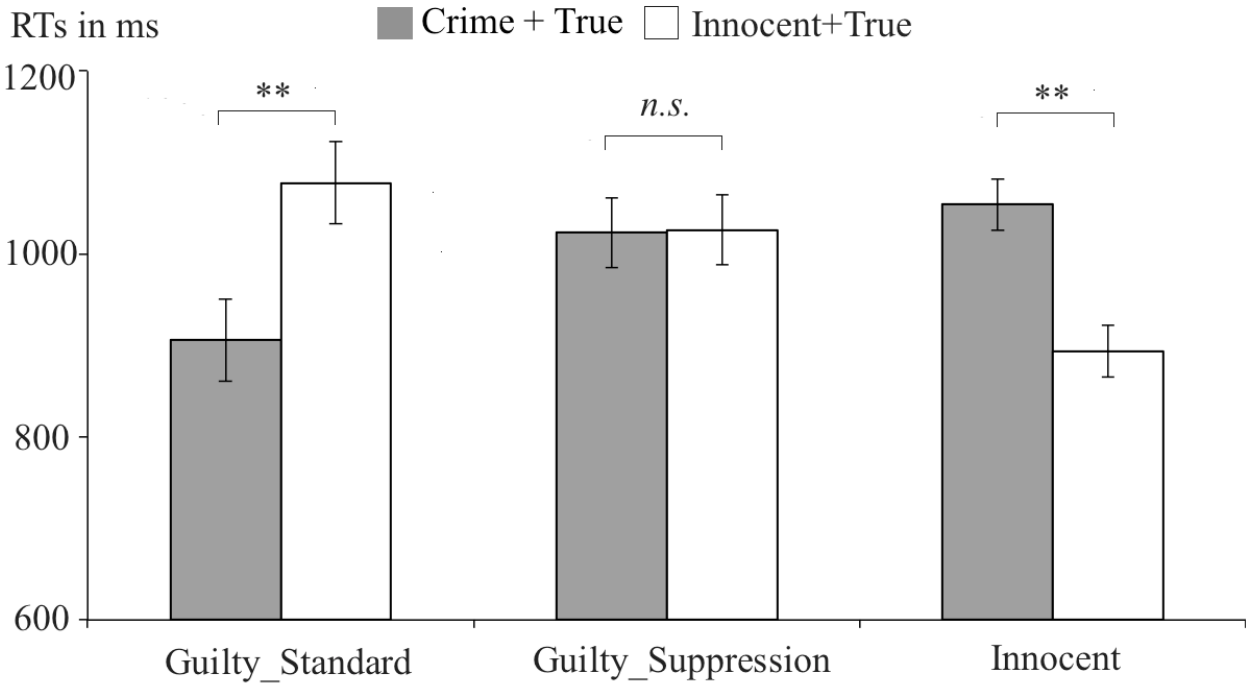


Because guilty/standard participants finished a CIT before the aIAT, this CIT may remind them of the crime and therefore artificially increase the aIAT effect. To address this concern, we compared the aIAT from the guilty/suppression group with comparable aIATs that were not preceded by CITs [Hu, Rosenfeld & Bodenhausen, 2012, (baseline aIATs) and Agosta & Sartori, 2013, (first aIAT administrations)]. Using these aIATs as a baseline, results still showed that suppression led to significantly reduced D-scores: Mean and 95% CIs (the non-overlapping 95% CIs indicate significant differences): for Suppression 0.13, [-0.03, 0.29] vs. 0.49 [0.40, 0.58] in Hu et al., 2012, N=64, vs. 0.58 [0.41, 0.73] in Agosta & Sartori, 2013, N=412. Thus, the effect of suppression on the aIAT is unlikely to be attributable to artificially increased aIAT scores when participants first complete the CIT.

RT Analyses: To better understand the reduction of D-scores and exclude concerns that participants distorted the aIAT results by intentionally slowing their responses, we analyzed RTs from the aIAT's double-classification blocks. A 3 (between-subject: guilty/ standard vs. guilty/suppression vs. innocent) by 2 (within-subject: congruent vs. incongruent blocks) mixed ANOVA showed that the group by block interaction was significant, $F(2,74)=19.04$, $p<.001$, $\eta_p^2=0.34$. Follow-up analyses showed that among innocent participants, the Innocence+True vs. Crime+True congruence effect was significant; Mean and 95% CIs: 893.67 ms [865.54, 921.80] vs. 1053.99 ms [1025.86, 1082.12], $t(25)=-5.59$, $p<.001$. Among guilty/standard participants, the Crime+True vs. Innocence+True congruence effect was also significant (905.63 ms [860.96, 950.30] vs. 1077.73 ms [1033.06, 1122.40], $t(25)=-3.78$, $p<.001$). In contrast, among guilty/suppression participants there was no Crime+True vs. Innocence+True congruence effect (1023.18 ms [985.11, 1061.25] vs. 1026.09 ms [988.03, 1064.16], $t(24)=-0.08$, $p>.90$, see Fig.4).

Employing the same Bayesian analysis procedure described for the P300, we found that the odds ratio favoring this null hypothesis (H_0) to the alternative hypothesis (H_1) is 6.48, $p(H_0|D)=0.87$.

Fig. 4: RTs from the Crime+True/Innocence+False and Innocence+True/Crime+False blocks in the aIAT. Error bars indicate 95% CIs. The Crime+True block is a congruent block for guilty yet an incongruent block for innocent participants; whereas the Innocence+True block is a congruent block for innocent but an incongruent block for guilty participants. ** $p<.001$.



Individual Classification Efficiency

The base-peak P300 successfully differentiated guilty/standard from innocent (AUCs=0.84, $p<.001$), as well as from guilty/suppression participants: (AUC=0.74, $p=.003$). However, the P300 could not differentiate between guilty/suppression and innocent participants, AUC=0.57, $p=.37$. Thus, suppression renders P300 ineffective in identifying guilty participants. However,

the LPN among guilty/suppression group still showed above-chance discrimination AUC=0.76, *p*=.001. Combining P300 and LPN in a peak-to-peak manner (i.e., P300-minus-LPN; Soskins et al., 2001) can discriminate guilty and innocent populations regardless of suppression or not, AUCs>0.70, *p*s<.01, see Table 1.

Table 1: Area under the curves (AUCs) and their 95% CIs from the Receiver Operating Characteristic (ROC) analyses.

Group	P300	LPN	P300 – LPN
Standard vs. Innocent	0.84 [0.72- 0.96]**	0.60 [0.45- 0.76]	0.80 [0.69- 0.92]**
Suppression vs. Innocent	0.57 [0.42- 0.73]	0.76 [.63- .89]**	0.73 [0.59- 0.87]**
Standard vs. Suppression	0.74 [0.60-0.88]**	0.63 [0.48 - 0.78]	0.56 [0.40 - 0.72]

Note: For P300- LPN combined measure, the LPN was subtracted from the P300 (P300 minus LPN). ** *p*<.01.

Post-experiment Questionnaires:

There were no differences between motivations to beat the test or nervousness during the lab crime ratings between the two guilty groups (*p*s>.12). Guilty/standard participants rated that the crime memories came to mind relatively automatically (3.62 ± 0.28 on a 0-6 scale, see SOM), but less automatically than in previous research (Bergström et al., 2013; obtained 3.90 ± 0.06 on a 1-4 scale). This discrepancy can be ascribed to different lab 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. In contrast, here 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 vs. simulation-based procedures could yield different levels of encoding depth of to-be-suppressed memories, which can account for differences in both suppression ERP effects and automaticity ratings between the two studies.

Exploratory Analyses

In addition to the hypothesis-driven analyses that are described above, exploratory analyses indicated that (1) suppression may have affected automatic aspects of aIAT performance more than controlled aspects, (2) guilty/suppression participants' aIAT performance could not be predicted by any of the measured ERP components, and (3) the P300 and LPN components were indeed orthogonal (for details, see the SOM).

Discussion

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

The amplitude of 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 to crime-relevant details provides direct neural evidence that

people can voluntarily terminate retrieving unwanted sensorimotor-rich memories. Critically, our guilty/standard group did not receive intentional retrieval instructions. Thus, the comparatively attenuated P300 in the guilty/suppression group is due to down-regulation of retrieval-related neural activity rather than up-regulation in the guilty/standard group, supporting the notion that inhibitory processes can directly override automatic retrieval (Anderson & Hanslmayr, 2014).

Despite their success at terminating recollection, guilty suppressors nevertheless revealed themselves via the enlarged LPNs. This LPN is dissociable from the recollection-sensitive P300s (Rugg et al., 1996), and may indicate response-monitoring processes (Johansson & Mecklinger, 2003). Here, guilty/suppression participants 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 guilty/suppression 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

visual memories can make them less visible in perceptual priming tasks (Gagnepain et al., 2014; Kim & Yi, 2013). Here, we obtained similar findings whereby 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, guilty suppressors behaved as if they had not experienced the lab crime. Together with previous research, it 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 its associated brain activity is problematic for neuroscience-based memory assessments. Nevertheless, suppression may leave its neural traces (LPN), suggesting that criminals using this countermeasure may still be identifiable with some memory detection protocols. Future research should test whether individual crime suppressors can be detected via fMRI, since suppression attempts engage the dorsolateral PFC (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.

Author Contributions:

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

References

Agosta, S., & Sartori, G. (2013). The autobiographical IAT: A review. *Frontiers in Psychology, 4*. doi:10.3389/Fpsyg.2013.00519

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 Science*. 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., 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

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

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. *Proc Natl Acad Sci U S A, 111*, E1310-1319. doi: 10.1073/pnas.1311468111

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. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Science, 6,* 299-305. doi:10.1016/S1364-6613(02)01923-X

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

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., 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

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

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

Acknowledgements: We thank Mike Anderson for providing the screening questionnaires. This study was supported by an APA Dissertation Research Award to X.Hu.