Journal of Experimental Psychology: General

Memory Suppression and Its Deficiency in Psychological Disorders: A Focused Meta-Analysis

Davide F. Stramaccia, Ann-Kristin Meyer, Katharina M. Rischer, Jonathan M. Fawcett, and Roland G. Benoit

Online First Publication, October 22, 2020. http://dx.doi.org/10.1037/xge0000971

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

http://dx.doi.org/10.1037/xge0000971

Memory Suppression and Its Deficiency in Psychological Disorders: A Focused Meta-Analysis

Davide F. Stramaccia and Ann-Kristin Meyer
Max Planck Institute for Human Cognitive and Brain Sciences

Katharina M. Rischer
Université du Luxembourg

Jonathan M. Fawcett
Memorial University of Newfoundland

Roland G. Benoit
Max Planck Institute for Human Cognitive and Brain Sciences

It is still debated whether suppressing the retrieval of unwanted memories causes forgetting and whether this constitutes a beneficial mechanism. To shed light on these 2 questions, we scrutinize the evidence for such suppression-induced forgetting (SIF) and examine whether it is deficient in psychological disorders characterized by intrusive thoughts. Specifically, we performed a focused meta-analysis of studies that have used the *think/no-think* procedure to test SIF in individuals either affected by psychological disorders or exhibiting high scores on related traits. Overall, across 96 effects from 25 studies, we found that avoiding retrieval leads to significant forgetting in healthy individuals, with a small to moderate effect size (0.28, 95% CI [0.14, 0.43]). Importantly, this effect was indeed larger than for more anxious (−0.21, 95% CI [−0.41, −0.02]) or depressed individuals (0.05, 95% CI [−0.19, 0.29])—though estimates for the healthy may be inflated by publication bias. In contrast, individuals with a stronger repressive coping style showed greater SIF (0.42, 95% CI [0.32, 0.52]). Furthermore, moderator analyses revealed that SIF varied with the exact suppression mechanism that participants were instructed to engage. For healthy individuals, the effect sizes were considerably larger when instructions induced specific mechanisms of direct retrieval suppression or thought substitution than when they were unspecific. These results suggest that intact suppression-induced forgetting is a hallmark of psychological well-being, and that inducing more specific suppression mechanisms fosters voluntary forgetting.

Keywords: suppression, involuntary retrieval, cognitive control, anxiety, depression

Forgetting is often regarded as a deficiency of our memory systems, where attempts to retain or retrieve information are met with failure. In particular, it has been argued to arise passively from either the temporal decay of the memory trace (Thorndike, 1913), interference from other memories that compete for retrieval (McGeoch, 1932; Underwood, 1957), or a change in context from initial encoding (Tulving, 1974).

However, under many circumstances forgetting can also be characterized as an adaptive force that shapes our memory, for instance by updating or discarding information that has become irrelevant—or even outright unwanted (Bjork, 1989; Bjork & Bjork, 1996; for reviews, see Fawcett & Hulbert, 2020; Nørby, 2015). Accumulating evidence suggests that such forgetting can be under intentional control: concerted attempts at preventing memories from entering awareness can subsequently make it more difficult to voluntarily retrieve these suppressed memories and eventually cause forgetting (Anderson & Green, 2001; Hertel & Calcaterra, 2005; see Anderson & Hanslmayr, 2014, for review). In essence, such suppression-induced forgetting (SIF; Anderson & Huddleston, 2012; Hertel & McDaniel, 2010) may serve the purpose of preventing our minds from being at the mercy of involuntary retrieval. The discarding of unwanted information may also more generally facilitate efficient cognition. For example, it supports response selection (Payne & Sekuler, 2014) and prevents excessive information intake that may otherwise increase uncertainty (Hertwig & Engel, 2016).

Davide F. Stramaccia and Ann-Kristin Meyer, Max Planck Research Group: Adaptive Memory, Max Planck Institute for Human Cognitive and Brain Sciences; Katharina M. Rischer, Department of Behavioural and Cognitive Sciences, Université du Luxembourg; Jonathan M. Fawcett, Department of Psychology, Memorial University of Newfoundland; Roland G. Benoit, Max Planck Research Group: Adaptive Memory, Max Planck Institute for Human Cognitive and Brain Sciences.

We thank Kerstin Flake and Heike Schmidt Duderstedt for their assistance in creating the figures. All data and R analysis scripts are available at the Open Science Framework (https://osf.io/f89ur/). The preprint of a preliminary version can be found at PsyArXiv (10.31234/osf.io/5wynm). Part of this work has been presented at the Psychologie und Gehirn Conference in May 2018. Roland G. Benoit, Davide F. Stramaccia, Katharina M. Rischer, and Ann-Kristin Meyer were funded by a Max Planck Research Group awarded to Roland G. Benoit. Jonathan M. Fawcett was funded by an NSERC Discovery Grant.

Correspondence concerning this article should be addressed to Davide F. Stramaccia or Roland G. Benoit. Max Planck Research Group: Adaptive Memory, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A, Leipzig 04103, Germany. E-mail: stramaccia@cbs.mpg.de or rbenoit@cbs.mpg.de

We here conducted a focused meta-analysis to scrutinize whether it is possible to foster forgetting *intentionally*. We were particularly interested in gauging whether such intentional forgetting may be a hallmark of psychological well-being (Benoit, Davies, & Anderson, 2016; Depue, Curran, & Banich, 2007; Engen & Anderson, 2018; Joormann, Hertel, LeMoult, & Gotlib, 2009; Visser, Lau-Zhu, Henson, & Holmes, 2018) and thus be deficient in people with disorders characterized by intrusive thoughts.

Intentional forgetting has been suggested to regulate our affective experience by preventing unwanted, affectively loaded memories from entering awareness. Moreover, unlike mere avoidance, it may reduce their accessibility—and possibly availability—in the long run and thus exert a persisting effect (Engen & Anderson, 2018). Intentional forgetting may also contribute to the reappraisal of experienced events and their emotional impact by overriding, or substituting, maladaptive responses with more favorable alternatives (Engen & Anderson, 2018; Hertel & Calcaterra, 2005).

Conversely, a deficiency in controlling one's memories and thoughts may be at the heart of several psychological disorders (e.g., Goschke, 2014; Hertel, 1997, 1998, 2007; McTeague, Goodkind, & Etkin, 2016). Perhaps most prominently, posttraumatic stress disorder (PTSD) is characterized by intrusive memories and unintentional reexperiencing (Brewin, 2014; Ehlers, Hackmann, & Michael, 2004; Hackmann, Ehlers, Speckens, & Clark, 2004). Indeed, this feature of PTSD has been recognized as one of its defining aspects in both the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM–5*; American Psychiatric Association, 2013) and the *International Statistical Classification of Diseases and Related Health Problems* (11th ed.; ICD; World Health Organization, 2018).

The intrusiveness of memories in PTSD may result from an impaired ability to keep unwanted memories at bay (Ehlers et al., 2004; Hackmann et al., 2004). Patients with this condition tend to seek help after intrusive memories are already strongly consolidated, thus highlighting the importance of understanding the retrieval processes that support the intrusions (Marks, Franklin, & Zoellner, 2018). Similarly, intrusive negative thoughts constitute central symptoms of other affective disorders such as anxiety (Kircanski, Johnson, Mateen, Bjork, & Gotlib, 2016) and depression (Kircanski, Joormann, & Gotlib, 2012). These intrusive thoughts have also been suggested to arise from the involuntary retrieval of previously experienced or imagined episodes (Iyadurai, Visser, et al., 2018; Visser et al., 2018).

To examine SIF in healthy and clinical populations, we meta-analyzed studies that have employed the *think/no-think* procedure (Anderson & Green, 2001).[1] In this procedure (see Figure 1), participants first learn to associate pairs of cues and targets (e.g., *TOMATO—VEST*), so that they can retrieve the target (*VEST*) upon presentation of its cue (e.g., *TOMATO*). Participants then enter the critical *think/no-think* phase, where they are shown a subset of the cues. For some of these cues, participants have to covertly rehearse the associated target (i.e., *recall* items). For other cues, participants need to actively prevent the associated target from coming to mind (i.e., *suppress* items). Each of those cues are presented several times, so to provide multiple opportunities for memory-control mechanisms to be deployed. A number of cues are not shown at all during this phase (i.e., *baseline* items), and serve to assess baseline memory performance in a following test phase. On that test, participants are instructed to recall each response (e.g., *VEST*) upon presentation of its specific cue (e.g., *TOMATO*), irrespective of previous instructions. Typically, participants are impaired at retrieving previously suppressed memories as indicated by worse memory accuracy for suppress than for baseline items. This finding of below-baseline memory accuracy is considered an index of SIF.

Though there has been accumulating evidence for SIF over the last 20 years (for review, see Anderson & Hanslmayr, 2014; Anderson & Huddleston, 2012; Wessel, Albers, Zandstra, & Heininga, 2020), including SIF-like effects in implicit/indirect memory tests (Gagnepain, Henson, & Anderson, 2014; Hertel, Large, Stück, & Levy, 2012; Hertel, Maydon, Ogilvie, & Mor, 2018; Wang, Luppi, Fawcett, & Anderson, 2019), this phenomenon has not universally been replicated (e.g., Algarabel, Luciano, & Martínez, 2006; Bergström, Velmans, de Fockert, & Richardson-Klavehn, 2007; Bulevich, Roediger, Balota, & Butler, 2006; Mecklinger, Parra, & Waldhauser, 2009; Wessel, Wetzels, Jelicic, & Merckelbach, 2005). A major goal of this analysis is thus to determine the statistical significance and magnitude of the SIF effect in healthy individuals. This is particularly important to also evaluate related deficits in clinical populations.

Some of the inconsistencies in the literature may reflect important study differences with respect to the exact mechanisms that people engaged to prevent unwanted retrieval. While initial studies were somewhat agnostic regarding the employed processes (e.g., Anderson & Green, 2001), there is now evidence for two specific suppression mechanisms. On one hand, people can prevent recall by stopping the retrieval process altogether (Benoit & Anderson, 2012; Bergström, de Fockert, & Richardson-Klavehn, 2009; Gagnepain et al., 2014). This mechanism, *direct retrieval suppression*, has been associated with an inhibitory top-down modulation of the hippocampus that originates from the right dorsolateral prefrontal cortex (Benoit & Anderson, 2012; Gagnepain et al., 2014). The other mechanism, *thought substitution*, requires participants to retrieve an alternative memory when faced with a cue to an unwanted memory. This substitute memory then occupies the limited focus of awareness and thus prevents the unwanted memory from coming to mind (Benoit & Anderson, 2012; Bergström et al., 2009; Hertel & Calcaterra, 2005). Thought substitution has been associated with memory selection processes supported by the left ventrolateral prefrontal cortex (Benoit & Anderson, 2012). Critically, both of these mechanisms have been shown to cause forgetting (Benoit & Anderson, 2012; Bergström et al., 2009; Hertel & Calcaterra, 2005). We will thus examine whether SIF in healthy individuals varies according to the induced suppression mechanism.

Turning to clinical populations, there is indeed evidence for impaired SIF, for example in PTSD (Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015; Sullivan et al., 2019; Wald-

[1] Note that there are also other experimental procedures that examine intentional forgetting. These include the *list-method directed forgetting* procedure (Bjork, 1970), which has also been linked to putative inhibitory mechanisms akin to those thought to cause SIF (Anderson & Hanslmayr, 2014; Bjork & Bjork, 1996; Bjork, 1989; cf. Sahakyan, Waldum, Benjamin, & Bickett, 2009). However, we focused on the *think/no-think* procedure because our aim was to assess motivated forgetting (a) at the stage of memory retrieval rather than encoding, and (b) directed at specific items in memory rather than lists (Anderson & Hanslmayr, 2014).

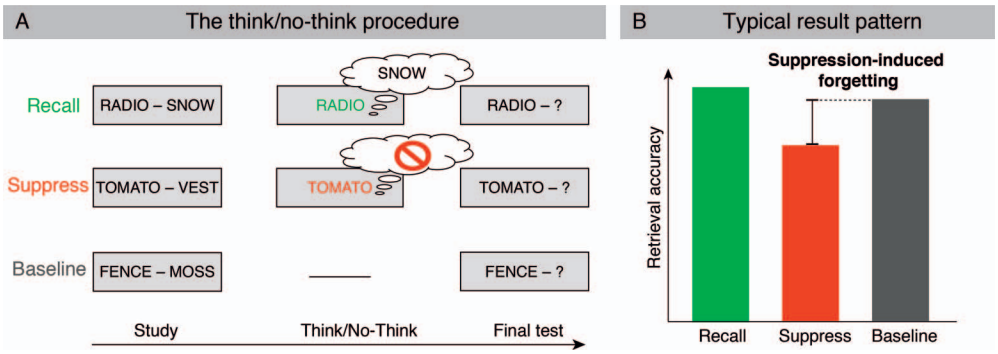

Figure 1. Panel A. Overview of the *think/no-think* procedure. In the initial study phase, participants encode associations of cues (e.g., RADIO) and targets (e.g., SNOW). They then enter the critical *think/no-think* phase, in which they repeatedly encounter most of the cues. For some of the cues (here for those presented in green), participants attempt to recall the associated targets (*recall* items). For other cues (here for those presented in red), their task is to prevent the associated target memory from coming to mind (*suppress* items). A third of the targets that they had also initially learned are not cued during this phase (*baseline* items). On a final test, participants are asked to remember all targets given their respective cues, irrespective of the previous instructions. Panel B. Typical retrieval accuracy on the final test. Participants are generally better or similarly capable at remembering *recall* than *baseline* targets. Critically, they are typically worse at retrieving previously suppressed than *baseline* targets. We refer to this finding as suppression-induced forgetting. See the online article for the color version of this figure.

hauser et al., 2018). However, the reliability of such a deficiency in clinical populations is still uncertain, because several studies did not directly observe impaired SIF (as compared with the respective healthy control group). Instead, these studies inferred memory control impairments from other between-groups differences that are less stringent indices of impaired intentional forgetting. These include better recall of *suppress* items (e.g., Hertel & Gerstle, 2003), impaired recall of *baseline* items (e.g., Hertel & Mahan, 2008), and different patterns of neural activation during the *think/no-think* phase as revealed by functional MRI (Sacchet et al., 2017).

To shed light on these issues, we meta-analyzed studies that compared clinical samples and subclinical samples (i.e., individuals displaying high scores on relevant clinical dimensions; see Search Strategy and Inclusion Criteria for details) with healthy controls on SIF as elicited by the *think/no-think* procedure. Specifically, we predicted a significant SIF effect for healthy control groups, as well as a significant difference between healthy control groups versus clinical and subclinical samples.

Furthermore, we explored the effects of a few important features that might influence the magnitude of SIF. First, we assessed the impact of providing different instructions that are either targeted at inducing specified mechanisms (i.e., *direct retrieval suppression* or *thought substitution*) or that leave it to the participants to prevent retrieval anyway they see fit (i.e., *unspecified* instructions). We hypothesized that participants would benefit from instructions that induce a specific mechanism. Intriguingly, it has been suggested that depressed individuals, whose cognitive control may be deficient, could particularly benefit from a mechanism like thought substitution that aids in avoiding unwanted retrieval by providing substitute memories (Hertel & Calcaterra, 2005).

Second, we examined whether the valence of the memories influences SIF, and whether this is especially the case for participants affected by (sub)clinical conditions. This is based on the idea that mood-congruent recall effects might modulate the effectiveness of memory control (Gaddy & Ingram, 2014; Matt, Vázquez, & Campbell, 1992). For instance, individuals with depression may be more prone to recall negative information, and therefore may also have a harder time suppressing it. Third, we tested whether more repetitions of a given *suppress* cue are associated with stronger SIF, as more *repetitions* provide more opportunities for successful suppression (as suggested by, e.g., Anderson & Green, 2001; Joormann et al., 2009). Fourth, we assessed the effects of presentation time for *suppress* cues. With longer presentation times, the suppression effort has to be sustained for a more extended period. This has recently been shown to cause more memory intrusions (van Schie & Anderson, 2017). We examine whether it also reduces SIF. Fifth, to inform future developments, we explored whether the effect size of SIF is sensitive to the type of *material* that had to be suppressed (i.e., *words* or *pictorial material*).

Method

Search Strategy and Inclusion Criteria

We sought to identify all studies that had used the *think/no-think* procedure to compare healthy groups with clinical or subclinical samples typically associated with cognitive control difficulties. We conducted our search in PubMed, Web of Science, and Google Scholar (on September 15, 2017),[2] using combinations of the following search terms: *think-no think* and/or *motivated forgetting*,

[2] In addition, as recommended by a reviewer, we also performed searches of the ERIC and Scopus databases with the "think/no-think" or "motivated forgetting" search terms in July 2019. However, these searches did not yield any additional paper that compared healthy with clinical or subclinical samples on SIF.

and disorders-related keywords: such as *thought control ability, impulsivity*, *anxiety*, *depression*, *dysphoria*, *ADHD* (attention-deficit/hyperactivity disorder), *OCD* (obsessive–compulsive disorder), *PTSD*, *schizophrenia*, *rumination*, *addiction*, *substance abuse*, *borderline*, *repressive coping*. (The term *suppression-induced forgetting* produced consistently redundant results and was dropped from the search strategy). Our literature search also included key terms related to questionnaires and tasks commonly associated with the broader literature on anxiety, depression, and thought control deficits. Specifically, these were the *STAI* (State–Trait Anxiety Inventory), *PANAS* (Positive and Negative Affect Schedule), *Beck Anxiety Inventory*, *Beck Depression Inventory*, *White Bear Suppression Inventory*, and the *Thought Control Ability questionnaire* (*TCAQ*; Luciano, Algarabel, Tomás, & Martínez, 2005). For exploratory purposes, we also included terms related to control and control deficits more broadly, that is, s*top-signal task*, *N-back*, *OSPAN* (Operation Span), *BIS-11* (Barratt Impulsiveness Scale), *Rumination Response Scale*, *go/no-go*, *Stroop*, and *flanker*. In addition, we consulted two recent review articles for additional references (Hulbert, Hirschstein, Brontë, & Broughton, 2018; Nørby, 2018), and included two studies that were published after the initial literature search had been completed (Noreen, Cooke, & Ridout, 2019; Waldhauser et al., 2018).

Finally, we attempted to identify pertinent studies that had not been published in peer-reviewed journals. Including such studies helps providing an overall SIF effect size estimate that is less influenced by publication bias (Thornton & Lee, 2000). In August, 2019, we therefore searched the *ProQuest* database for otherwise unpublished dissertation projects using the terms "think/no-think" and "motivated forgetting." In addition, in August, 2019, we e-mailed the corresponding authors of relevant publications. These included the authors of the clinical think/no-think studies that we had identified in the literature search described above.

We further extended this call to corresponding authors of other articles on SIF (i.e., those not studying SIF in clinical populations) and of other articles on the related topics of *retrieval-induced forgetting* and *list-method directed forgetting* (as identified through *PubMed* and *Web of Science*; $N = 56$). This procedure led to the inclusion of a doctoral dissertation that contained otherwise unpublished *think/no-think* data that matched our inclusion criteria (see below; Ryckman, 2015).

Moreover, if necessary, we further asked the authors of the included think/no-think studies for all the information required to compute effect sizes or, alternatively, for the respective data sets so that we could extract them ourselves. We thus received additional information for six studies (Depue, Burgess, Willcutt, Ruzic, & Banich, 2010; Hertel & Gerstle, 2003; Hertel & Mahan, 2008; Hertel & McDaniel, 2010; Stephens, Braid, & Hertel, 2013; Wessel et al., 2005).

For all data that we had received directly from the respective authors, we used the newly obtained descriptive statistics instead of those extracted from the articles (many of which required direct extraction from the figures as described below). This procedure also provided the necessary information that allowed us to include the article by Stephens, Braid, and Hertel (2013) into the study pool.

The inclusion procedure for the retrieved studies is summarized in Figure 2, following the recommendation of Moher, Liberati, Tetzlaff, and Altman (2009). We included all studies that compared at least one clinical sample to a healthy control group. We also included, as subclinical samples, studies with groups of participants that scored high on questionnaires of clinical relevance (i.e., *BDI*, *STAI*, and *RRS*), or studies that split their participants into subclinical and control groups based on such questionnaires. We included only studies that used the *think/no-think* procedure; that reported at least one test outcome pertaining to episodic

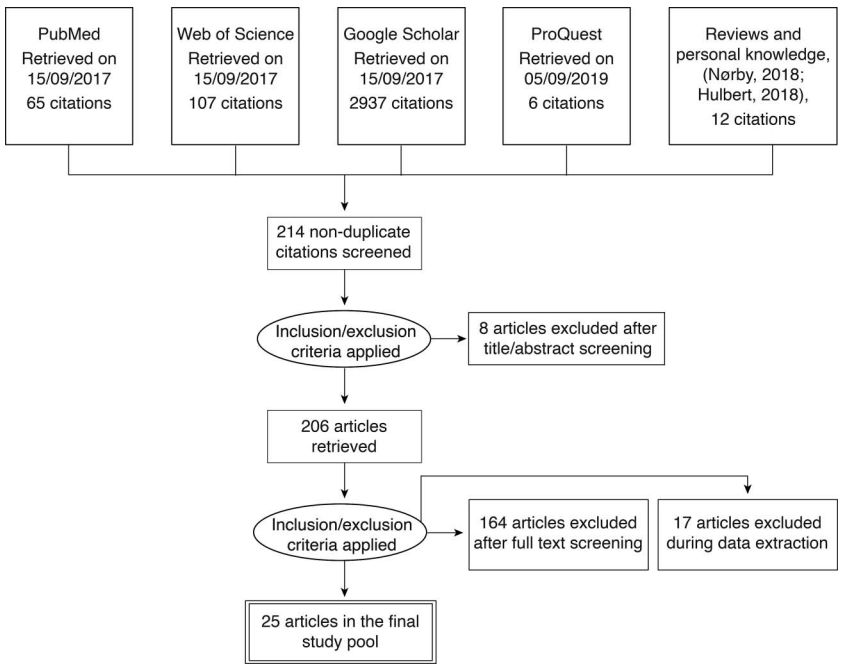

Figure 2. Schematic overview of the literature search and inclusion process.

memory performance; that reported sufficient data for the meta-analysis either in text, figures, supplementary material; or that were made available in response to our request.

Data Extraction

In total, the search yielded 214 unique entries, of which 25 entered our quantitative analysis (see Table 1). These were coded by three of the authors (DFS, KR, and AK-M). DFS and A-KM had extensive knowledge of the SIF literature and practical expertise with the *think/no-think* procedure; KR had previous experience with literature search for meta-analyses.

In a first step, DFS and KR jointly recorded recall performance (for "same probe" tests as opposed to "independent probe" tests; see section Significant Suppression-Induced Forgetting in Healthy Adults) for *baseline* and *suppress* items of each group, as well as five potential moderators of the effect (see below). They also coded the nature of the clinical or subclinical condition.

In a second step, to ensure the reliability of the data extraction, a third author (A-KM) independently coded all of the information. A-KM and DSF then examined their interrater agreement and reached full consensus with respect to the moderators and clinical condition. For many of the included studies, the critical mean values and measures of dispersion were only provided in plots (see Table 1). In these cases, they manually extracted these values using WebPlotDigitizer (Rohatgi, 2017), which has been shown to yield high intercoder reliability (Drevon, Fursa, & Malcolm, 2017).

Indeed, our two sets of coding also only yielded minor differences. We thus obtained high interrater reliability on those measure, in terms of high criterion-referenced reliability, a case of intraclass correlation coefficient (*ICC*; McGraw & Wong, 1996) suitable for comparing our coding. Indeed, the lowest *ICC* was 0.98, and all coefficients were significantly different from 0 (all $p < .001$). We thus deemed it appropriate to reach a consensus by averaging the two respective sets of values.

Several studies reported multiple, nonindependent measures of SIF. These included retrieval accuracy on different test formats and multiple ways of rating the quality of the retrieved memories. Similarly, some studies employed *within-subject* manipulations of, for example, the number of repetitions during the *think/no-think* phase or the valence of the *suppress* items. They therefore provided multiple estimates of SIF (i.e., one for each level of the within-subject manipulation). In general, we included all the non-independent SIF measures. This was always the case for effects related to moderators of interest (e.g., SIF from different sets of *suppress* items each characterized by a different emotional *valence*). For studies that employed experimental manipulations other than those identified as moderators of interest (e.g., homograph or nonhomograph stimuli in Hertel & McDaniel, 2010), we coded all the nonindependent SIF measures unless the authors had only provided aggregate data across levels of the manipulation.[3] Similarly, whenever task features were manipulated *between-groups*, we included all independent SIF effects, as long as it was possible to distinguish performance of healthy from that of clinical and subclinical participants. For example, when different samples of participants suppressed either positive or negative items, we included both of the resulting independent effect sizes. For two studies (Kim, Oh, Kim, Sim, & Lee, 2013; Kim, Yi, Yang, & Lee, 2007), we could not obtain any dispersion measures for memory performance. We estimated the missing standard deviations (see Higgins & Green, 2011, p. 485) by calculating the respective means of the standard deviations of the other included studies, weighted by their respective sample sizes (for this procedure, we excluded the few instances where outcomes were not reported in percentage form; Catarino et al., 2015; Küpper, Benoit, Dalgleish, & Anderson, 2014).

In addition, we coded for the five potential moderators of SIF. First, we coded the nature of the *instructions* given to participants to prevent retrieval (*direct retrieval suppression*, *thought substitution*, or *unspecified*). One study had different participants assigned to either *unspecified* or *thought substitution* instructions (Noreen & Ridout, 2016b), but did not provide separated SIF results as a function of both, *instructions* and *group*. For each group, we therefore took the SIF effects combined across the two instruction conditions and marked them as *unspecified*.

Second, we coded the *valence* of the stimulus material (for the suppress targets only) as either neutral, positive, negative, or mixed (i.e., when the only reported effect sizes were combined across different valence levels). When studies comprehensively reported SIF for different valence categories assigned to the same participants (e.g., for neutral, negative, and positive memories in Marzi, Regina, & Righi, 2014; neutral and negative in Sacchet et al., 2017; Zhang, Xie, Liu, & Luo, 2016), we generally included the effect size related to each condition, where available. Finally, for one study (Dieler, Herrmann, & Fallgatter, 2014) we coded SIF for negative items only, because its analysis of group differences (low vs. high anxiety) did not include neutral items.

Third, we coded the *repetitions* of *suppress* items, that is, the number of times that participants encountered each cue in the *think/no-think* phase. One study reported a SIF effect averaged across two conditions with two and eight repetitions (Noreen & Ridout, 2016b). We here coded the average (five) as the number of repetitions associated with that effect size, as we could not obtain the data set to disentangle the two.

Fourth, we coded the duration for which cues remained on the screen during the *think/no-think* phase, and, fifth, the *material* of stimuli that participants had to suppress (i.e., *words* or *pictorial material*). One study (Stephens et al., 2013) that had examined the recall of autobiographical memories was also coded as *words*.

Statistical Analysis

Our main focus was twofold: assessing the statistical significance and magnitude of SIF in healthy individuals and determining whether SIF is indeed reduced in (sub)clinical samples characterized by intrusive thoughts and deficits of cognitive

[3] With this approach, we assess the robustness of SIF across many different manipulations and means of quantifying SIF. In Appendix A, we provide two complementary random-effects models (one for the healthy and one for the (sub)clinical samples) that only include the single independent effect size of each study that constitutes the condition or measurement most typically used to assess SIF in the extant literature (for justifications of this approach, see Card, 2015, pp. 192–193; Cooper et al., 2019, p. 282; Higgins and Green, 2011, Chapter 3; Lipsey and Wilson, 2001, p. 125). These models thus provide an estimate of the presumably strongest manipulations (e.g., the greatest rather than fewer suppression repetitions). (These models further allow for a comparison with our initial preprint available at PsyArXiv doi: 10.31234/osf.io/5wynm).

Table 1
Studies/Effect Sizes Included in the Meta-Analyses

Figure 3 ID	Study	StudyID	SampleID	Comments	Language	Instructions	Stimuli		Duration(s)	Repetitions	*N*	DV	(Sub)clinical sample	Cluster	Group	SIF (SMCC)
							Material	Valence								
[1]	Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015	1	1		English	Direct RS	Pictures	Negative	3	10	18	Correct ID	PTSD	Anxiety	Emotional disorders	−0.45
[2]	Catarino et al., 2015	1	1		English	Direct RS	Pictures	Negative	3	10	18	Correct gist	PTSD	Anxiety	Emotional disorders	0.08
[3]	Catarino et al., 2015	1	1		English	Direct RS	Pictures	Negative	3	10	18	Correct details	PTSD	Anxiety	Emotional disorders	−0.07
[4]	Catarino et al., 2015	1	2		English	Direct RS	Pictures	Negative	3	10	18	Correct ID		Healthy	Healthy	0.25
[5]	Catarino et al., 2015	1	2		English	Direct RS	Pictures	Negative	3	10	18	Correct gist		Healthy	Healthy	0.70
[6]	Catarino et al., 2015	1	2		English	Direct RS	Pictures	Negative	3	10	18	Correct details		Healthy	Healthy	0.68
[7]	Depue, Burgess, Willcutt, Ruzic, & Banich, 2010	2	3		English	Unspecified	Pictures	Negative	3.5	12	16	Retrieval ACC		Healthy		0.73
[8]	Depue et al., 2010	2	4		English	Unspecified	Pictures	Negative	3.5	12	16	Retrieval ACC	ADHD	Mixed		0.05
[9]	Dieler, Herrmann, & Fallgatter, 2014	3	5		German	Unspecified	Pictures	Negative	4	12	36	Retrieval ACC	High trait anxiety	Anxiety	Emotional disorders	−0.44
[10]	Dieler et al., 2014	3	6		German	Unspecified	Pictures	Negative	4	12	35	Retrieval ACC		Healthy	Healthy	0.10
[11]	Diwadkar et al., 2017	4	7		Italian	Unspecified	Words	Negative	4	8	10	Retrieval ACC	GAD	Anxiety	Emotional disorders	−0.42
[12]	Diwadkar et al., 2017	4	8		Italian	Unspecified	Words	Negative	4	8	10	Retrieval ACC		Healthy	Healthy	−0.46
[13]	Fawcett et al., 2015	5	9		English	Direct RS	Words	Negative	3	12	24	Retrieval ACC		Healthy	Healthy	0.75
[14]	Fawcett et al., 2015	5	10		English	Direct RS	Words	Negative	3	12	24	Retrieval ACC	High rumination	Depression	Emotional disorders	0.27
[15]	Hertel & Gerstle, 2003	6	11		English	Unspecified	Words	Neutral	3	8	31	Retrieval ACC		Healthy	Healthy	−0.36
[16]	Hertel & Gerstle, 2003	6	11		English	Unspecified	Words	Neutral	3	16	31	Retrieval ACC		Healthy	Healthy	−0.23
[17]	Hertel & Gerstle, 2003	6	12		English	Unspecified	Words	Neutral	3	8	31	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.19
[18]	Hertel & Gerstle, 2003	6	12		English	Unspecified	Words	Neutral	3	16	31	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.02
[19]	Hertel & Mahan, 2008	7	13	Related pair	English	Unspecified	Words	Neutral	3	12	18	Retrieval ACC		Healthy	Healthy	0.19
[20]	Hertel & Mahan, 2008	7	14	Unrelated pair	English	Unspecified	Words	Neutral	3	12	18	Retrieval ACC		Healthy	Healthy	0.17
[21]	Hertel & Mahan, 2008	7	15	Related pair	English	Unspecified	Words	Neutral	3	12	18	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.08
[22]	Hertel & Mahan, 2008	7	16	Unrelated pair	English	Unspecified	Words	Neutral	3	12	18	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.08
[23]	Hertel & McDaniel, 2010	8	17	Aided supp.; homographic cues	English	Thought sub.	Words	Negative	3	12	18	Retrieval ACC		Healthy		0.67
[24]	Hertel & McDaniel, 2010	8	17	Aided supp.; non-homographic cues	English	Thought sub.	Words	Negative	3	12	18	Retrieval ACC		Healthy		0.38
[25]	Hertel & McDaniel, 2010	8	18	Unaided supp.; homographic cues	English	Unspecified	Words	Negative	3	12	18	Retrieval ACC		Healthy		0.27
[26]	Hertel & McDaniel, 2010	8	18	Unaided supp.; non-homographic cues	English	Unspecified	Words	Negative	3	12	18	Retrieval ACC		Healthy		0.52
[27]	Hertel & McDaniel, 2010	8	19	Aided supp.; homographic cues	English	Thought sub.	Words	Negative	3	12	18	Retrieval ACC	High repression	Repression		0.81
[28]	Hertel & McDaniel, 2010	8	19	Aided supp.; non-homographic cues	English	Thought sub.	Words	Negative	3	12	18	Retrieval ACC	High repression	Repression		0.87
[29]	Hertel & McDaniel, 2010	8	20	Unaided supp.; homographic cues	English	Unspecified	Words	Negative	3	12	18	Retrieval ACC	High repression	Repression		0.29
[30]	Hertel & McDaniel, 2010	8	20	Unaided supp.; non-homographic cues	English	Unspecified	Words	Negative	3	12	18	Retrieval ACC	High repression	Repression		0.31
[31]	Joormann, Hertel, LeMoult, & Gotlib, 2009	9	21		English	Unspecified	Words	Negative	3	12	15	Retrieval ACC		Healthy	Healthy	0.68
[32]	Joormann et al., 2009	9	22		English	Unspecified	Words	Negative	3	12	15	Retrieval ACC	Depression	Depression	Emotional disorders	−0.26

(*table continues*)

Table 1 (*continued*)

Figure 3 ID	Study	StudyID	SampleID	Comments	Language	Instructions	Stimuli		Duration(s)	Repetitions	*N*	DV	(Sub)clinical sample	Cluster	Group	SIF (SMCC)
							Material	Valence								
[33]	Joormann et al., 2009	9	23	Positive substitutes	English	Thought sub.	Words	Negative	3	12	15	Retrieval ACC		Healthy	Healthy	0.94
[34]	Joormann et al., 2009	9	24	Negative substitutes	English	Thought sub.	Words	Negative	3	12	15	Retrieval ACC		Healthy	Healthy	0.38
[35]	Joormann et al., 2009	9	25	Positive substitutes	English	Thought sub.	Words	Negative	3	12	15	Retrieval ACC	Depression	Depression	Emotional disorders	0.76
[36]	Joormann et al., 2009	9	26	Negative substitutes	English	Thought sub.	Words	Negative	3	12	15	Retrieval ACC	Depression	Depression	Emotional disorders	0.75
[37]	Joormann, Hertel, Brozovich, & Gotlib, 2005	10	27		English	Unspecified	Words	Negative	4	12	18	Retrieval ACC		Healthy	Healthy	0.28
[38]	Joormann et al., 2005	10	28		English	Unspecified	Words	Positive	4	12	18	Retrieval ACC		Healthy	Healthy	0.00
[39]	Joormann et al., 2005	10	29		English	Unspecified	Words	Negative	4	12	18	Retrieval ACC	Depression	Depression	Emotional disorders	0.50
[40]	Joormann et al., 2005	10	30		English	Unspecified	Words	Positive	4	12	18	Retrieval ACC	Depression	Depression	Emotional disorders	0.09
[41]	Kim, Kim, Sim, & Lee, 2013	11	31		Korean	Unspecified	Pictures	Negative	4	8	14	Correct hits		Healthy		0.66
[42]	Kim et al., 2013	11	32		Korean	Unspecified	Pictures	Negative	4	8	13	Correct hits	High repression	Repression		0.44
[43]	Kim, Yi, Yang, & Lee, 2007	12	33		Korean	Unspecified	Words	Negative	4	16	46	Retrieval ACC		Healthy		0.02
[44]	Kim et al., 2007	12	33		Korean	Unspecified	Words	Neutral	4	16	46	Retrieval ACC		Healthy		0.42
[45]	Kim et al., 2007	12	33		Korean	Unspecified	Words	Negative	4	8	46	Retrieval ACC		Healthy		−0.03
[46]	Kim et al., 2007	12	33		Korean	Unspecified	Words	Neutral	4	8	46	Retrieval ACC		Healthy		0.36
[47]	Kim et al., 2007	12	34		Korean	Unspecified	Words	Negative	4	16	29	Retrieval ACC	High repression	Repression		0.44
[48]	Kim et al., 2007	12	34		Korean	Unspecified	Words	Neutral	4	16	29	Retrieval ACC	High repression	Repression		0.40
[49]	Kim et al., 2007	12	34		Korean	Unspecified	Words	Negative	4	8	29	Retrieval ACC	High repression	Repression		0.33
[50]	Kim et al., 2007	12	34		Korean	Unspecified	Words	Neutral	4	8	29	Retrieval ACC	High repression	Repression		0.36
[51]	Küpper, Benoit, Dalgleish, & Anderson, 2014	13	35		English	Direct RS	Pictures	Negative	3	10	12	Correct ID		Healthy		0.67
[52]	Küpper et al., 2014	13	35		English	Direct RS	Pictures	Negative	3	10	12	Correct gist		Healthy		0.31
[53]	Küpper et al., 2014	13	35		English	Direct RS	Pictures	Negative	3	10	12	Correct details		Healthy		1.93
[54]	Küpper et al., 2014	13	36		English	Direct RS	Pictures	Negative	3	10	12	Correct ID	Low thought control	Mixed		0.50
[55]	Küpper et al., 2014	13	36		English	Direct RS	Pictures	Negative	3	10	12	Correct gist	Low thought control	Mixed		0.45
[56]	Küpper et al., 2014	13	36		English	Direct RS	Pictures	Negative	3	10	12	Correct details	Low thought control	Mixed		0.64
[57]	Marzi, Regina, & Righi, 2014	14	37		Italian	Direct RS	Pictures	Negative	3.5	5	15	Retrieval ACC	High trait anxiety	Anxiety	Emotional disorders	0.06
[58]	Marzi et al., 2014	14	37		Italian	Direct RS	Pictures	Neutral	3.5	5	15	Retrieval ACC	High trait anxiety	Anxiety	Emotional disorders	0.44
[59]	Marzi et al., 2014	14	37		Italian	Direct RS	Pictures	Positive	3.5	5	15	Retrieval ACC	High trait anxiety	Anxiety	Emotional disorders	0.62
[60]	Marzi et al., 2014	14	38		Italian	Direct RS	Pictures	Negative	3.5	5	15	Retrieval ACC		Healthy	Healthy	1.31
[61]	Marzi et al., 2014	14	38		Italian	Direct RS	Pictures	Neutral	3.5	5	15	Retrieval ACC		Healthy	Healthy	0.81
[62]	Marzi et al., 2014	14	38		Italian	Direct RS	Pictures	Positive	3.5	5	15	Retrieval ACC		Healthy	Healthy	0.91
[63]	Nemeth et al., 2014	15	39		Hungarian	Unspecified	Words	Neutral	3.5	16	36	Retrieval ACC		Healthy		0.07
[64]	Nemeth et al., 2014	15	40		Hungarian	Unspecified	Words	Neutral	3.5	16	36	Retrieval ACC	Alcohol abuse	Mixed		−0.46
[65]	Noreen & Ridout, 2016b	16	41	Aided/unaided supp.	English	Unspecified	Words	Neutral	3	5	36	Retrieval ACC		Healthy	Healthy	0.35
[66]	Noreen & Ridout, 2016b	16	42	Aided/unaided supp.	English	Unspecified	Words	Neutral	3	5	36	Retrieval ACC	Dysphoria	Depression	Emotional disorders	0.24
[67]	Noreen & Ridout, 2016a	17	43	Aided supp.	English	Thought sub.	Words	Mixed	3	8	18	Retrieval ACC		Healthy	Healthy	0.45
[68]	Noreen & Ridout, 2016a	17	44	Unaided supp.	English	Unspecified	Words	Mixed	3	8	18	Retrieval ACC		Healthy	Healthy	−0.70
[69]	Noreen & Ridout, 2016a	17	45	Aided supp.	English	Thought sub.	Words	Mixed	3	8	18	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.82
[70]	Noreen & Ridout, 2016a	17	46	Unaided supp.	English	Unspecified	Words	Mixed	3	8	18	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.85
[71]	Noreen, Cooke, & Ridout, 2019	18	47	Aided supp.	English	Thought sub.	Words	Mixed	3	8	61	Retrieval ACC		Healthy	Healthy	0.33
[72]	Noreen et al., 2019	18	48	Aided supp.	English	Thought sub.	Words	Mixed	3	8	60	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.48
[73]	Ryckman, 2015	19	49	Experiment 2	English	Unspecified	Pictures	Neutral	5	12	17	Retrieval ACC		Healthy		0.24
[74]	Ryckman, 2015	19	50	Experiment 2	English	Unspecified	Pictures	Neutral	5	12	19	Retrieval ACC	High neuroticism	Mixed		−0.20

(*table continues*)

Table 1 (*continued*)

Figure 3 ID	Study	StudyID	SampleID	Comments	Language	Instructions	Stimuli		Duration(s)	Repetitions	*N*	DV	(Sub)clinical sample	Cluster	Group	SIF (SMCC)
							Material	Valence								
[75]	Ryckman, 2015	19	51	Experiment 3 and 4	English	Unspecified	Words	Negative	4.5	16	101	Retrieval ACC		Healthy		−0.17
[76]	Ryckman, 2015	19	52	Experiment 3 and 4	English	Unspecified	Words	Negative	4.5	16	102	Retrieval ACC	High neuroticism	Mixed		0.09
[77]	Sacchet et al., 2017	20	53		English	Unspecified	Words	Negative	3	12	16	Retrieval ACC		Healthy	Healthy	0.15
[78]	Sacchet et al., 2017	20	53		English	Unspecified	Words	Neutral	3	12	16	Retrieval ACC		Healthy	Healthy	0.23
[79]	Sacchet et al., 2017	20	54		English	Unspecified	Words	Negative	3	12	16	Retrieval ACC	Depression	Depression	Emotional disorders	0.14
[80]	Sacchet et al., 2017	20	54		English	Unspecified	Words	Neutral	3	12	16	Retrieval ACC	Depression	Depression	Emotional disorders	0.29
[81]	Salamé & Danion, 2007	21	55		French	Unspecified	Words	Neutral	3	8	24	Retrieval ACC		Healthy		0.06
[82]	Salamé & Danion, 2007	21	55		French	Unspecified	Words	Neutral	3	16	24	Retrieval ACC		Healthy		0.06
[83]	Salamé & Danion, 2007	21	56		French	Unspecified	Words	Neutral	3	8	23	Retrieval ACC	Schizophrenia	Mixed		0.28
[84]	Salamé & Danion, 2007	21	56		French	Unspecified	Words	Neutral	3	16	23	Retrieval ACC	Schizophrenia	Mixed		0.56
[85]	Stephens, Braid, & Hertel, 2013	22	57		English	Unspecified	Words	Neutral	4	12	30	Retrieval ACC		Healthy	Healthy	0.04
[86]	Stephens et al., 2013	22	57		English	Unspecified	Words	Mixed	4	12	30	Retrieval ACC		Healthy	Healthy	−0.12
[87]	Stephens et al., 2013	22	58		English	Unspecified	Words	Neutral	4	12	30	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.08
[88]	Stephens et al., 2013	22	58		English	Unspecified	Words	Mixed	4	12	30	Retrieval ACC	Dysphoria	Depression	Emotional disorders	−0.05
[89]	Waldhauser et al., 2018	23	59		German/English/ Native	Direct RS	Pictures	Neutral	2	12	11	Correct hits	PTSD	Anxiety	Emotional disorders	−0.24
[90]	Waldhauser et al., 2018	23	60		German/English/ Native	Direct RS	Pictures	Neutral	2	12	13	Correct hits		Healthy	Healthy	0.52
[91]	Wessel, Wetzels, Jelicic, & Merckelbach, 2005	24	61		Dutch	Unspecified	Words	Neutral	3	16	33	Retrieval ACC		Healthy		0.27
[92]	Wessel et al., 2005	24	62		Dutch	Unspecified	Words	Neutral	3	16	35	Retrieval ACC	High dissociation	Mixed		0.26
[93]	Zhang, Xie, Liu, & Luo, 2016	25	63		Chinese	Direct RS	Pictures	Negative	3	10	25	Retrieval ACC		Healthy	Healthy	0.52
[94]	Zhang et al., 2016	25	63		Chinese	Direct RS	Pictures	Neutral	3	10	25	Retrieval ACC		Healthy	Healthy	0.63
[95]	Zhang et al., 2016	25	64		Chinese	Direct RS	Pictures	Negative	3	10	25	Retrieval ACC	Dysphoria	Depression	Emotional disorders	0.19
[96]	Zhang et al., 2016	25	64		Chinese	Direct RS	Pictures	Neutral	3	10	25	Retrieval ACC	Dysphoria	Depression	Emotional disorders	0.81

Note. Figure 3 ID corresponds to respective references axis labels in the forest plot (Figure 3). StudyID and SampleID are the random effects labels that group effect sizes together based on whether they belong to the same study and experimental group, respectively. Language specifies the language of experimental material and/or more generally the language in which the experiment was carried out. Instructions indicate whether the respective study aimed at inducing a specific suppression mechanism. Instructions as well as material, valence, duration, and repetitions were included as potential moderators. The dependent variable identifies the respective test used to quantify suppression-induced forgetting. (Sub)clinical sample identifies the specific clinical or subclinical condition investigated in each study, and *Cluster* the broader (sub)clinical set it was assigned to (e.g., anxiety). *Group* indicates more broadly whether the effect size reported in the row stems from a healthy or (sub)clinical group with an emotional disorder. *N* indicates the size of the respective samples. SIF (suppression induced forgetting) indicates the within-group effects (standardized mean change with change score standardization, SMCC) of the (sub)clinical and healthy groups. Supp. = suppression (comments column); Direct RS = direct retrieval suppression; Thought sub. = thought substitution (instructions column); *N* = sample size; DV = dependent variable; Correct ID = correct identification; Retrieval ACC = retrieval accuracy (DV column).

control. We therefore computed a series of meta-analyses in R 3.6.1 (R Development Core Team, 2008) with the package *metafor* 2.1.0 (Viechtbauer, 2010).

We first clustered studies based on clinical and subclinical conditions. Specifically, we grouped the clinical samples with respect to the psychiatric taxonomy of the *DSM* (5th ed., American Psychiatric Association, 2013; i.e., anxiety, depression; note that we grouped PTSD with anxiety, given that this is an often co-occurring feature of this disorder) or the similarity of their defining characteristic (i.e., high repressive coping style). We added the subclinical samples according to their relatedness along the psychopathological continuum (e.g., we combined depressed mood with major depressive disorder; Table 1).

We thus identified a *depression cluster* (20 effect sizes from 11 studies, including major depressive disorder, dysphoria, and rumination, $N = 324$ (sub)clinical participants), an *anxiety cluster* (nine effect sizes from five studies, including high trait anxiety, generalized anxiety disorder [GAD] and PTSD, $N = 90$ (sub)clinical participants), and a *repression cluster* (nine effect sizes from three studies, including high repressive coping, $N = 78$ (sub)clinical participants). Repression, unlike the other (sub)clinical *clusters*, has previously been linked to a stronger ability to prevent retrieval, and we thus expected greater SIF for this *cluster* (Hertel & McDaniel, 2010).

We assigned the remaining effect sizes to a *mixed cluster* (10 effect sizes from six studies, with $N = 243$ (sub)clinical participants). This *cluster* included one study each on alcohol abuse, ADHD, schizophrenia, low thought control ability (as measured by the *TCAQ*; Luciano et al., 2005; greater scores on the *TCAQ* are negatively associated with both anxiety and depression as well as obsessive–compulsive disorder; Williams et al., 2010), dissociative disorders (as measured by the Dissociative Experiences Scale, *DES*; Bernstein & Putnam, 1986), and high neuroticism (as measured by the neuroticism subscale of the Eysenck Personality Questionnaire—Revised, *EPQ-R*; Eysenck, Eysenck, & Barrett, 1985). The high heterogeneity of samples included in the *mixed cluster* hinders meaningful comparisons with the other, more clinically defined *clusters*. We therefore only examine it on its own and refrain from any comparison. Finally, the effect sizes of all control groups were combined in one *healthy cluster* (48 effect sizes, $N = 687$ healthy participants).

We computed all the effect sizes as the standardized mean change score between *baseline* and *suppress* items (as implemented in the *escalc* function; measure set to *SMCC*), using the extracted means and standard deviations. However, this method requires an estimate of the correlation between *baseline* and *suppress* items, which was not reported in the surveyed literature. We thus estimated the Spearman correlation as $r = .3$, based on data from our group and on the studies for which we were able to obtain the respective data sets (Hertel & Gerstle, 2003; Hertel & Mahan, 2008; Hertel & McDaniel, 2010; Stephens et al., 2013; Wessel et al., 2005). The original correlation coefficient was retained for these data sets. (Note that additional sensitivity analyses, using alternative correlation coefficients of $r = .1$ and .6, yielded the same conclusions with respect to our main hypothesis.) Accordingly, only results obtained with an assumed correlation of .3 will be reported. Two studies (Catarino et al., 2015; Küpper et al., 2014) had employed three fairly different measures of SIF. For these, we used correlation coefficients of three unpublished data sets from our group for which we had used the same measures.

We then performed a random-effects meta-regression (Hedges & Olkin, 1985) of the SIF effect sizes, grouped by *cluster* ($N = 1{,}588$ participants; 96 effect sizes from 25 studies). To account for the correlation between some of the effect sizes, we used a multilevel random-effects (MLRE) model (Konstantopoulos, 2011), with random effects (intercepts) for both the sample (i.e., sampleID) and study (i.e., studyID) from which the data were derived, with the former nested within the latter. In an earlier iteration of the analysis, we had fitted four-level models including an additional random effect for each of the individual effect sizes. However, there was no advantage in this approach because that variable explained little to no variance, and ultimately did not yield any difference. We therefore removed it in favor of a simpler model structure.

We then estimated robust confidence intervals for the effect sizes with the *robust.rma.mv* function, with study identifier (*studyID*) as the clustering variable used for constructing the sandwich estimator (see Hedges, Tipton, & Johnson, 2010). This approach allowed us to account for violations in the independence assumptions due to multiple effect sizes arising from the same participants, which were frequent in our data set; and, more generally, for the correlation between effects within each study. We used this approach for all of the following analyses.

To assess whether this analysis was more informative than a simpler random-effects meta-analysis of all effect sizes irrespective of any distinction by *cluster*, we used Akaike's information criterion (*AIC*; Akaike, 1998) with small-sample correction (*AICc*), transformed to conditional probabilities for each model (Wagenmakers & Farrell, 2004). The resulting *AIC* weights (*AICw*) thus provide evidence for the relative fit of the two compared models to the data (note that all *AICw* for a set of models sum up to 1).

We computed *AICc* and *AICw* using the *fitstats* (from the *metafor* package) and *akaike.weights* (from the *qpcR* package; Spiess, 2018) functions in *R*. In general, we fitted our models using restricted maximum likelihood (REML) as opposed to the maximum likelihood (ML) method due to ML's bias in variance components' estimates. However, AIC is not suitable for comparing sets of models that have been fitted with REML and differ in their fixed effects structures. The model comparisons were thus based on models that were refitted using ML. We report the best fitting model.

We complemented the meta-analyses of the healthy versus (sub)clinical samples with a series of further moderator analyses. We performed separate analyses for each of the five moderators (*instructions*, *valence*, *repetitions*, *duration*, and *material*). These analyses were carried out on a reduced sample of studies, combining the *anxiety* and *depression clusters* in a single *emotional disorders* group (e.g., Goodwin, 2015). These choices were due to the relatively small pool of studies that contributed to each level of the moderators and to achieve a reasonable homogeneity of the included clinical samples. The moderator analyses thus do not include the *repression cluster*, due to its—expected—opposite effect on intentional forgetting (Hertel & McDaniel, 2010), and the *mixed cluster*, due to the diversity of the samples' (sub)clinical characteristics. Overall, this approach, with *group* (*healthy* vs.

emotional disorders) rather than *cluster*, thus served to increase the power of the analyses while maintaining theoretical consistency.

For each moderator analysis, we compared the *AICw* of a model that included a moderator**group* interaction with that of a moderator+*group* model that only included main effects. We report only the best fitting model, or, in cases where models performed similarly, the simpler one. We had planned to carry out these five moderator analyses for theoretical reasons, and they were further motivated by the high heterogeneity consistently observed in our models. Therefore, we expected that the chosen moderators might explain part of this heterogeneity.

For all the reported models, we evaluated heterogeneity across samples by calculating the 95% prediction interval (PI; IntHout, Ioannidis, Rovers, & Goeman, 2016). The PI indexes the range of effects expected from new samples similar to those included in the analysis. Heterogeneity was also tested with Cochran's *Q* (Cochran, 1954), where a significant outcome rejects the null hypothesis that all the included studies evaluated the same effect (*QE* was used for models that included moderators). We further examined I^2, which indicates how much of the overall variation across studies is due to heterogeneity as opposed to mere chance (Higgins, Thompson, Deeks, & Altman, 2003). Specifically, we used a generalized form (Nakagawa & Santos, 2012) that allowed us to quantify such proportions separately for higher-level (I^2_{studyID}, for *studyID*) and nested (I^2_{sampleID}, for *sampleID*) random factors.

Meta-analyses are susceptible to publication bias, that is, the inflation or otherwise distortion of effect size estimates due to selective reporting of favorable study outcomes (Thornton & Lee, 2000) and other forms of questionable research practices (Renkewitz & Keiner, 2019). In particular, in the context of the present meta-analyses, there could be a bias for reporting experiments that yielded a significant SIF effect for the healthy individuals. Alternatively, there could be a bias for publishing studies that did not find significant SIF in the (sub)clinical groups.

Ideally, this problem would be mitigated by including all unpublished studies, assuming that they were not published for exactly these biases. However, we only retrieved three such experiments (Ryckman, 2015).

Therefore, to further gauge these biases, we used contour-enhanced funnel plots to display each study's effect size against its precision as indexed by the standard error (Peters, Sutton, Jones, Abrams, & Rushton, 2008). These plots are centered at zero, and display areas of statistical significance. This, in turn, allows for easier visual detection of publication bias due to exclusion of studies that yielded nonsignificant results. We plotted effect sizes separately for healthy individuals and (sub)clinical groups.

We then used Egger's regression test (e.g., Egger, Davey Smith, Schneider, & Minder, 1997; Peters, Sutton, Jones, Abrams, & Rushton, 2006) to formally assess funnel plot asymmetry as an indicator of publication bias (with $p < .1$ as the critical value, following the recommendation of Egger et al., 1997). Because this test is not yet implemented for MLRE models,[4] we performed it by reestimating each model (healthy individuals or (sub)clinical groups) with the inclusion of a moderator coding for the standard error of the effect sizes. A significant deviation from zero in the intercept of this meta-regression would indicate that the relationship between precision and size of the studies is asymmetrical, and thus biased (Sterne & Egger, 2005).

Because the sensitivity of meta-analytic estimates is also vulnerable to outliers in the study pool, we also evaluated the included studies for influential cases, based on Cook's distance (*cooks.distance.rma.mv*, clustered by *studyID*). This is a leave-one-out diagnostic measure (available in *metafor*) that is suitable for data sets with a multilevel structure and for the robust estimation of confidence intervals (Cook & Weisberg, 1982; Viechtbauer & Cheung, 2010). We thus compared the model that was most informative in respect to our hypotheses—the MLRE meta-regression with studies grouped by *cluster*—fitted with and without studies that appeared to be highly influential as indicated by a Cook's distance greater than 1 (Hair, Black, Babin, Anderson, & Tatham, 1998).

Results

Suppression-Induced Forgetting in Healthy Versus (Sub)Clinical Samples

First, we assessed whether a model using the *cluster* moderator (i.e., *healthy*, *anxiety*, *depression*, *repression*, and *mixed*) would provide a better fit to the data than a simpler model with effect sizes classified as just *healthy* or (sub)*clinical*, or than the basic *null* model without any moderator. Indeed, this was the case, with the *cluster* model ($AICw_{cluster} = .72$) being approximately 2.8 times more plausible than the *simple* model ($AICw_{simple} = .26$) and 36 times more plausible than the *null* model ($AICw_{null} = .02$). The *cluster* model also displayed the lowest overall heterogeneity, $I^2 = 67.85\%$. Overall, these results support using the *cluster* model as the benchmark for testing our two main hypotheses.

Importantly, the overall effect of the moderator *cluster* was significant, $F(4, 20) = 18.93$, $p < .001$. There was a significant small-to-moderate SIF effect of 0.28, 95% CI [0.14, 0.43], 95% PI [−.38, .95], $p < .001$ for the *cluster* of all *healthy* samples. By comparison, the *anxiety cluster* displayed a small significant effect in the opposite direction (indicating significantly higher recall of *suppress* than *baseline* items), with an estimate of −0.21, 95% CI [−0.41, −0.02], 95% PI [−0.89, 0.47], $p = .036$. The *depression cluster* did not show evidence in support of any effect, with an estimate of 0.05, 95% CI [−0.19, 0.29], 95% PI [−0.64, 0.74], $p = .662$. The effect for the *mixed cluster* was in the direction of SIF, but failed to reach significance, with an estimate of 0.17, 95% CI [−0.09, 0.43], 95% PI [−0.53, 0.87], $p = .188$. However, the *repression cluster* yielded a significant effect, with an estimate of 0.42, 95% CI [0.32, 0.52], 95% PI [−0.23, 1.08], $p < .001$. As shown in the preceding paragraph, there was a high amount of heterogeneity, which was also significant, $QE_{(91)} = 251.92$, $p < .001$.

Comparisons of the individual *clusters* displayed a significant difference for the *healthy* compared to the *anxiety clus-*

[4] For the same reason, we could not adjust for publication bias using the trim-and-fill procedure (Duval & Tweedie, 2000). However, in Appendix A, we additionally perform this procedure for the simpler random-effects models that are based on only the single effect sizes from each study that are derived from the most typical measures and manipulations. Appendix B provides further simple random-effects models using fill-and-trim correction based on the average effect sizes from each study. Note that we apply these methods irrespective of the non-significant Egger's test of the main analysis.

ter, −0.50, 95% CI [−0.65, −0.34], $p < .001$, and the *depression cluster*, −0.23, 95% CI [−0.45, −0.01], $p = .039$. Therefore, the results corroborate our hypothesis of significant SIF in healthy individuals and of impaired SIF in clinical samples associated with *emotional disorders* (see Figure 3). Finally, there was significantly greater SIF for the small sample of effect sizes arising from the *repression cluster*, compared with the *healthy cluster*, with a difference of 0.14, 95% CI [0.05, 0.22], $p = .003$.

Because the *cluster* model still exhibited high heterogeneity, we next investigated whether influential studies or our choice of correlation coefficients might have driven the results. We thus followed an identical approach to the one described in the previous section. Here, Cook's distance indicated one particularly influential study (Marzi et al., 2014) that exceeded the set threshold of 1. Refitting the *cluster* model without these effect sizes (90 effect sizes rather than 96) yielded comparable results. Furthermore, heterogeneity was still significant, $QE_{(83)} = 219.09$, $p < .001$, and not lower than before, $I^2 = 65.06\%$, although more evenly spread between the random factors ($I^2_{studyID} = 37.70\%$, $I^2_{sampleID} = 27.36\%$). In light of these results, we chose to keep this study in the pool for the subsequent analyses.

Moderators Analysis

Though the meta-analysis provided evidence for significant SIF in the general population, the included studies varied widely in the mechanism that individuals were instructed to adopt to prevent unwanted retrieval, as well as with respect to the *material*, *valence*, *repetitions*, and *duration* of the *suppress items*. For each moderator model, we first checked whether a model including *group* (*healthy* vs. *emotional disorders*) and the respective *moderator* provided a better fit to the data with the interaction of these factors or, more simply, with their linear combination only. Because the *emotional disorders* group combined the *depression* and *anxiety clusters*, we first verified that it also showed lower SIF than the *healthy* group. This was the case, with a significant difference of −0.29, 95% CI [−0.48, −0.10], $F(1, 14) = 10.9781$, $p = .005$.

Instructed mechanism. There was a slight advantage for the *instructions*group* compared with the *instructions + group* model, with the former ($AICw_{instructions*group} = 0.76$) being approximately 3.2 times more plausible than the latter ($AICw_{instructions + group} = 0.24$). Therefore, we reported the *instructions*group* model (see Table 2).

The *instructions* moderator significantly differentiated between memory control mechanisms, though only in *healthy* individuals.

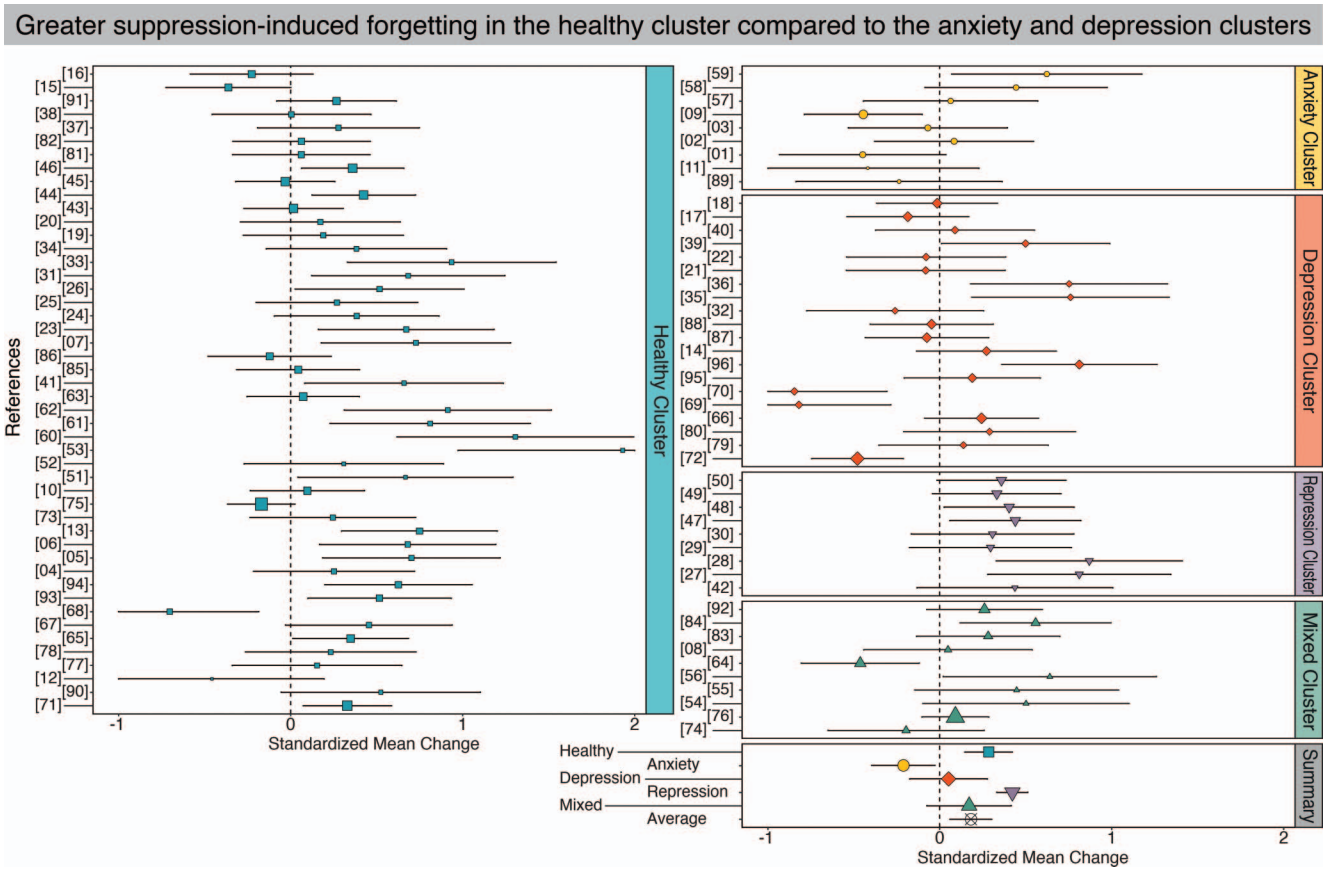

Figure 3. Forest plot of all effect sizes grouped by *cluster*. Standardized mean changes with change score standardization and 95% confidence interval, as a function of clinical *cluster*. References point to the corresponding Figure 3 IDs in Table 1. Symbols for individual effect sizes are sized proportionally to the respective sample sizes. Symbols at the bottom display the meta-analytic effect sizes from the meta-regression models and the overall effect size from the random-effects model (not sized proportionally to sample sizes). See the online article for the color version of this figure.

Table 2
Instructed Mechanism Moderator Meta-Regression

Coefficients	Estimate (*SE*)	*t*	*p*	95% CI [lower CI, upper CI]
Intercept	0.66 (0.11)	6.22	<.001	[0.42, 0.90]
Instructed mechanism				
Thought substitution	−0.08 (0.14)	−0.56	.589	[−0.38, 0.23]
Unspecified	−0.68 (0.16)	−4.15	.002	[−1.04, −0.31]
Group				
Emotional disorders	−0.49 (0.14)	−3.44	.006	[−0.81, −0.17]
Interaction				
Thought Substitution × Emotional Disorders	−0.10 (0.40)	−0.26	.803	[−0.98, 0.78]
Unspecified × Emotional Disorders	0.38 (0.18)	2.11	.062	[−0.02, 0.78]

Note. Intercept Corresponds to the estimated average effect for levels *direct retrieval suppression* and *healthy* of the *instructed mechanism* and *group* factors, respectively. The omnibus test of moderators was significant, $F(5, 10) = 8.20$, $p = .003$, whereas the *Instructed Mechanism* × *Group* interaction was not, $F(2, 10) = 2.29$, $p = .152$. Heterogeneity was large, $I^2 = 61.43$ ($I^2_{\text{studyID}} = 49.24$, $I^2_{\text{sampleID}} = 12.19$), and significant, $QE_{(52)} = 117.99$, $p < .001$.

For these samples, *direct retrieval suppression* displayed a medium SIF effect of 0.66, with 95% CI [0.42, 0.90], 95% PI [−0.02, 1.34], $p < .001$. The SIF effects for *thought substitution* was also significant, with 0.59, 95% CI [0.40, 0.77], 95% PI [−0.08, 1.25], $p < .001$. However, the SIF effect of *unspecified instructions* was not significant, with −0.02, (95% CI [−0.29, 0.26], 95% PI [−0.71, 0.68], $p = .902$).

In the *((sub)clinical* individuals, by contrast, none of the instructed mechanisms were significant. D*irect retrieval suppression* displayed a small nonsignificant SIF effect of 0.17, with 95% CI [−0.17, 0.51], 95% PI [−0.55, 0.90], $p = .288$. SIF for *thought substitution* was also not significant, with an estimate of −0.01, 95% CI [−0.85, 0.84], 95% PI [−1.06, 1.53], $p = .987$. *Unspecified instructions* yielded a numerically reversed, albeit also not significant, SIF effect of −0.13, 95% CI [−0.41, 0.16], 95% PI [−0.83, 0.57], $p = .344$).

Notably, for healthy individuals, the SIF effect for direct retrieval suppression was significantly higher than the one for unspecified instructions, with a difference of 0.68, (95% CI [0.31, 1.04], $p = .002$). This was also the case for the comparison of thought substitution and unspecified instructions, with a difference of 0.60, 95% CI [0.28, 0.92], $p = .002$. The effects for direct retrieval suppression and thought substitution were quite similar to each other, with a negligible nonsignificant difference of 0.08, (95% CI [−0.23, 0.38], $p = .589$) in favor of the former. The results thus corroborate the importance of instructing a specific mechanism to elicit SIF (Figure 4A). However, none of these comparisons were significant for the *emotional disorders cluster* (all $p < .165$).

Material. The *material* × *group* model was very similar to the *material* + *group* model, with the former ($AICw_{material \times group} = 0.46$) being about as plausible as the latter ($AICw_{material + group} = 0.54$). We thus decided to focus on the simpler model for the *material* moderator (see Table 3).

For the *healthy group, r*esults revealed a significant SIF for *pictorial material*, with a small-to-moderate effect of 0.42, (95% CI [0.09, 0.75], 95% PI [−0.39, 1.23], $p = .016$), and a trend for *words*, with a small effect of 0.21, (95% CI [−0.02, 0.44], 95% PI [−0.57, 0.99], $p = .067$). (The same analysis on all healthy control samples from the study pool, i.e., not just those from the studies on emotional disorders, yielded significant SIF of 0.17, 95% CI [0.02, 0.31], 95% PI [−0.49, 0.82], $p = .026$.) SIF for *pictorial material* was twice as large as for *words*, though not significantly different, 0.21, 95% CI [−0.18, 0.61], $p = .264$. In addition, as expected, the average effect was significantly smaller for the *emotional disorders group* (see Table 3).

Valence. The *valence*group* model did not improve on the simpler *valence* + *group* model, with the former ($AICw_{valence^*group} = 0.10$) in fact being nine times less plausible than the latter ($AICw_{valence + group} = 0.90$). We thus decided to focus on the model without an interaction for the *valence* moderator (see Table 4).

In terms of different *valences* of *suppress* stimuli, in *healthy groups*, the *neutral valence* condition displayed a small yet statistically significant SIF effect of 0.33, with 95% CI [0.04, 0.62], 95% PI [−0.55, 1.27], $p = .029$. Likewise, the *negative valence* condition was significant with an estimate of 0.29, 95% CI [0.18, 0.56], 95% PI [−0.48, 1.06], $p = .039$. The *positive valence*

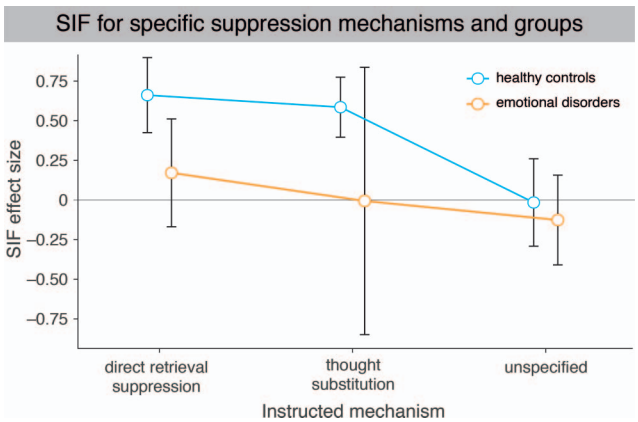

Figure 4. Average suppression-induced forgetting (SIF) effect size for *instructed mechanisms*, separately for the *healthy* and *emotional disorders* groups. Standardized mean changes with change score standardization and 95% confidence interval, as a function of *group* and *instructions*. Blue circles denote values for *healthy groups*, whereas orange circles indicate values for (sub)clinical samples with *emotional disorders*. See the online article for the color version of this figure.

Table 3
Material Moderator Meta-Regression

Coefficients	Estimate (*SE*)	*t*	*p*	95% CI [lower CI, upper CI]
Intercept	0.21 (0.10)	2.00	.067	[−0.02, 0.44]
Stimuli				
Pictures	0.21 (0.18)	1.17	.264	[−0.18, 0.61]
Group				
Emotional disorders	−0.29 (0.09)	−3.2	.007	[−0.49, −0.10]

Note. Intercept corresponds to the estimated average effect for levels *words* and *healthy* of the *material* and *group* factors, respectively. The omnibus test of moderators was significant, $F(2,v13) = 5.68$, $p = .017$. Heterogeneity was large, $I^2 = 70.21$ ($I^2_{studyID} = 41.95$, $I^2_{sampleID} = 28.26$), and significant, $QE_{(55)} = 159.73$, $p < .001$.

condition was relatively similar to the *neutral condition*, but not significant, with an estimate of 0.36, 95% CI [−0.19, 0.91], 95% PI [−0.55, 1.27], $p = .177$. Instead, the *mixed condition* exhibited a negligible SIF effect, which was not significant, with an estimate of 0.08, 95% CI [−0.22, 0.39], 95% PI [−0.70, 0.87], $p = .557$. As before, the average effect was significantly smaller for the *emotional disorders* group (see Table 4). However, none of the comparisons were significant (all $p \geq .154$).

Repetitions. In terms of *AICc*, the *repetitions* × *group* ($AICw_{repetitions \times group} = 0.36$) model was slightly inferior to the *repetitions* + *group* model ($AICw_{repetitions + group} = 0.64$). Therefore, we report the simpler model without an interaction for the *repetitions* moderator (see Table 5). However, the main effect of *repetitions* was not significant, $F(1, 13) = 0.59$, $p = .456$, providing no evidence that the amount of *repetitions* (within the included range) influences SIF.

Duration. The *duration* × *group* model ($AICw_{duration \times group} = 0.40$) was approximately 1.5 times worse than the *duration* + *group* model ($AICw_{duration + group} = 0.60$). Therefore, we report the model without an interaction for the *duration* moderator (see Table 6). The main effect of *duration* was not significant, $F(1, 13) = 1.14$, $p = .306$, and thus we did not find evidence for a role of *duration* of the *suppress* stimuli on SIF.

Publication Bias Analysis

For the effect sizes arising from *healthy individuals* only, the contour-enhanced funnel plot exhibited some apparent degree of right-skewness (Figure 5a). However, many of the effect sizes fell within the areas of nonsignificance, suggesting a lighter bias. Indeed, Egger's regression for this set of effect sizes was not significant, with $p = .182$, and thus did not provide evidence for an asymmetrical relationship between the precision and effect size of the studies.

For the effect sizes arising from (*sub)clinical samples*, points in the contour-enhanced funnel plot were quite evenly scattered around the meta-analytic effect size estimate (Figure 5b). As before, Egger's regression for this set of effect sizes was not significant, with $p = .409$.

Discussion

In a series of meta-analyses, we set to examine two questions: (a) whether preventing retrieval can cause forgetting and (b) whether such suppression is deficient in individuals with psychological disorders that are characterized by intrusive thoughts. We therefore focused on studies employing the *think/no-think* procedure that compared the SIF effect between healthy control groups and relevant clinical and subclinical samples. In the following, we will first discuss the replicability, effect size, mechanisms, possible causes, and moderators of SIF in the general population. We will then turn to the evidence for impaired SIF in patient populations and discuss the implications of the results for theorizing about memory suppression as a beneficial coping mechanism.

Significant Suppression-Induced Forgetting in Healthy Adults

Our analyses of the healthy individuals demonstrated a significant albeit small-to-moderate effect size. Inspection of the contour-

Table 4
Valence Moderator Meta-Regression

Coefficients	Estimate (*SE*)	*t*	*p*	95% CI [lower CI, upper CI]
Intercept	0.33 (0.13)	2.51	.029	[0.04, 0.62]
Valence				
Mixed	−0.25 (0.16)	−1.53	.154	[−0.60, 0.11]
Negative	−0.04 (0.13)	−0.32	.754	[−0.33, 0.25]
Positive	0.03 (0.21)	0.14	.889	[−0.44, 0.50]
Group				
Emotional disorders	−0.29 (0.10)	−2.95	.013	[−0.51, −0.07]

Note. Intercept corresponds to the estimated average effect for levels *neutral* and *healthy* of the *valence* and *group* factors, respectively. The omnibus test of moderators was not significant, $F(4, 11) = 2.57$, $p = .097$. Heterogeneity was large, $I^2 = 67.72$ ($I^2_{studyID} = 35.15$, $I^2_{sampleID} = 32.57$), and significant, $QE_{(53)} = 146.75$, $p < .001$.

Table 5
Repetitions Moderator Meta-Regression

Coefficients	Estimate (*SE*)	*t*	*p*	95% CI [lower CI, upper CI]
Intercept	0.16 (0.21)	0.77	.453	[−0.29, 0.62]
Repetitions	0.01 (0.01)	0.77	.456	[−0.02, 0.04]
Group				
Emotional disorders	−0.29 (0.09)	−3.20	.007	[−0.49, −0.10]

Note. Intercept corresponds to the estimated average effect for the extrapolated value of *repetitions* = 0. The omnibus test of moderators was significant, $F(2, 13) = 5.79$, $p = .016$. Heterogeneity was large, $I^2 = 71.41$ ($I^2_{\text{studyID}} = 43.76$, $I^2_{\text{sampleID}} = 27.65$), and significant, $QE_{(55)} = 171.70$, $p < .001$.

enhanced funnel plots suggested some degree of publication bias, though this was not corroborated by Egger's regression test (though see Caveats). We further examined publication bias for SIF by performing trim-and-fill corrections on two complementary simple random-effects models (Appendices A and B; though see section Caveats). One model was based on the average effect sizes of each study (see Appendix B). For this model, the penalty of the trim-and-fill procedure was—surprisingly—greater when including unpublished studies that themselves had reported null results or even significant reversals than when just including published studies. The former analysis yielded an effect at the trend level only, though the latter remained significant. The other model (see Appendix A) was based on only the single independent effect sizes of each study that constituted the conditions or measurement of SIF that are most common in the literature. It thus presumably provides an estimate of the strongest manipulations (e.g., the greatest number of suppression repetitions; Anderson & Green, 2001). This effect remained significant following trim-and-fill. Overall, the results thus corroborate that memory suppression can induce forgetting.

The healthy participants in the included studies were typically matched to the respective clinical sample on demographic measures. As a corollary, individuals of the control groups were closer to a community sample than what is usually realized in psychology experiments. We therefore suggest that the results of this analysis may be fairly generalizable to a wider population. For the same reason, however, we might have underestimated the upper boundary of the effect size that could be achieved by high functioning, young adults.

At the same time, it is also possible that we underestimated its lower boundary, because the control individuals were typically selected to lack the clinical features that had been of interest to the respective study. Therefore, the control groups may be mentally healthier—and therefore more adept at controlling unwanted memories—than a random sample of the general population.

Across the meta-analyses, we also examined features that might influence SIF in healthy individuals and in the emotional disorders of anxiety and depression. Overall, we did not find a substantial contribution of *valence*, *repetitions*, or *duration* of suppression attempts toward the magnitude of SIF. These null findings may be a consequence of the overall little variance in levels of the moderators (in particular for *repetitions* and *duration*; see also section Caveats). However, we found that SIF was numerically greater for *pictorial material* than for *words*. This finding is encouraging, given that suppressing pictures, with their greater complexity, may be more similar to suppressing autobiographical memories and thus possess higher ecological validity (e.g., Stephens et al., 2013). We note, though, that there are also a number of methodological differences between studies using words versus pictures. For example, the latter tend to use more fine grained dependent variables rather than simple binary accuracy measures, which may make the former more sensitive for assessing subtle forgetting effects. However, SIF was neither significant for pictorial material nor for words in the (sub)clinical samples.

Most notably, in healthy samples, there were substantial differences in SIF due to task *instructions*. Though all included studies formally used the *think/no-think* procedure to assess intentional forgetting, they differed with respect to whether they left it to the participants to find possible solutions to prevent retrieval or whether they prescribed a specific mechanism (either *direct retrieval suppression* or *thought substitution*). In healthy participants, SIF was significantly greater under *direct retrieval suppression* and *thought substitution* instructions compared to *unspecified* instructions.

These results thus clearly indicate that it is essential for future *think/no-think* studies to provide specific instructions. Indeed, Her-

Table 6
Duration Moderator Meta-Regression

Coefficients	Estimate (*SE*)	*t*	*p*	95% CI [lower CI, upper CI]
Intercept	0.71 (0.41)	1.71	.111	[−0.19, 1.60]
Duration	−0.13 (0.13)	−1.07	.306	[−0.41, 0.14]
Group				
Emotional disorders	−0.29 (0.09)	−3.21	.007	[−0.49, −0.10]

Note. Intercept corresponds to the estimated average effect for the extrapolated value of *duration* = 0. The omnibus test of moderators was significant, $F(2, 13) = 6.24$, $p = .013$. Heterogeneity was large, $I^2 = 71.25$ ($I^2_{\text{studyID}} = 43.91$, $I^2_{\text{sampleID}} = 27.34$), and significant, $QE_{(55)} = 171.08$, $p < .001$.

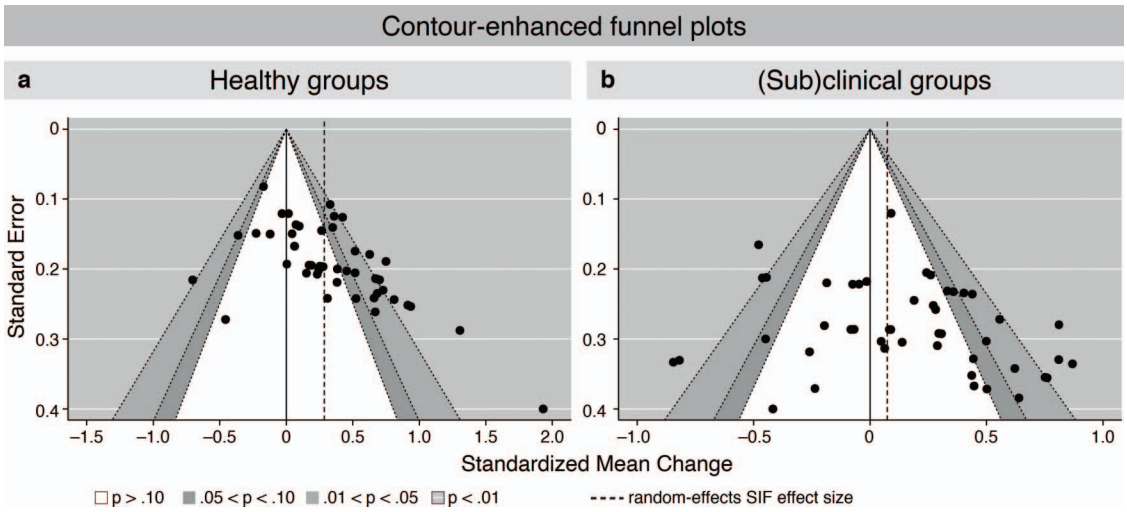

Figure 5. Contour-enhanced funnel plots of *healthy* (panel a) and *(sub)clinical* (panel b) samples displaying individual effect sizes (black circles). The black dashed lines indicate the estimated effect sizes. As indicated in the legend, each background color indicates a different area of significance within which a study may fall into: white for $p > .10$, dark gray for $.05 < 0 < .10$, gray for $.01 < p < .05$, and light gray for $p < .01$. See the online article for the color version of this figure.

tel and Calcaterra (2005) had previously provided evidence for stronger SIF when participants were using *thought substitution* rather than following *unspecified* instructions. In general, *unspecified* instructions require participants to first find possible solutions to prevent involuntary retrieval. They may also lead participants to alternate between a multitude of suppression mechanisms throughout the procedure. Both of these may diminish the efficacy of suppression and thus weaken the degree of SIF.

Suppression mechanisms may not only differ in the underlying neuro-cognitive processes (e.g., Benoit & Anderson, 2012; Bergström et al., 2009), but also in the manner that they induce subsequent forgetting. The prominent inhibitory account of memory control suggests that suppression attempts lead to the recruitment of inhibitory processes that directly target and weaken the avoided memory trace (Anderson & Hanslmayr, 2014; Detre, Natarajan, Gershman, & Norman, 2013). As a consequence, the targeted memories should become less available in subsequent retrieval attempts and, accordingly, should only have a reduced influence on later thoughts (e.g., Wang et al., 2019). However, in many situations, preventing retrieval may also hinder subsequent recall of the unwanted memory by noninhibitory processes such as associative interference (Hertel & Calcaterra, 2005; Verde, 2013; Racsmány, Conway, Keresztes, & Krajcsi, 2012; Tomlinson, Huber, Rieth, & Davelaar, 2009).

Interference may particularly contribute to forgetting following thought substitution, which likely strengthens the association between the cue (e.g., *TOMATO*) and the alternate thought or memory that participants had retrieved (e.g., *CLOWN*) to prevent the unwanted target memory from coming to mind (e.g., *VEST*). In fact, some authors have suggested that thought substitution may be predominantly achieved through interference (Bergström et al., 2009; Hertel & McDaniel, 2010; Racsmány, Conway, Keresztes, & Krajcsi, 2012; cf., Benoit & Anderson, 2012; see Belli, 2011, for extensive discussion), similarly to retroactive interference in the *A-B, A-C* paradigm (Briggs, 1954). Though recent evidence suggests that inhibition also plays a role in this paradigm (Hulbert & Anderson, 2020).

Behavioral evidence for inhibitory versus noninhibitory accounts of SIF is provided by studies that employed an independent probe procedure to assess forgetting (Anderson & Green, 2001; Bergström et al., 2009). In these studies, the suppressed memories are also probed with a new cue that has a strong preexperimental association with the memory (e.g., its category; *DRESS—V* for *VEST*). This test thus probes the memory while circumventing its association with the original cue (e.g., *TOMATO*). As such, SIF on an independent probe test is unlikely to be caused by associative interference (cf. Racsmány et al., 2012). Instead, it is more likely caused by a weakened representation of the suppressed memory, consistent with an inhibitory account of memory control.

Of the included studies, very few used an independent probe test. It is thus difficult to gauge the evidence for inhibitory versus noninhibitory forgetting. However, an exploratory analysis of the four studies that included such a test (comprising eight effect sizes across healthy and (sub)clinical samples) revealed a trend for a small effect only, SMCC = 0.18, 95% CI [−0.0356, 0.402], p = .088. However, all of these studies provided *unspecified* instructions or used a *thought substitution* procedure, and particularly the latter has only inconsistently been associated with inhibitory forgetting (Benoit & Anderson, 2012; Bergström et al., 2009). The efficacy of thought substitution in inducing inhibitory forgetting may hinge on the exact choice of the substitute memories and their relatedness to the unwanted memories (Benoit & Anderson, 2012; Hertel & McDaniel, 2010; Norman, Newman, & Detre, 2007). These exploratory results should encourage future meta-analytical treatments of SIF as measured by independent probes (https://osf.io/hmctu), with particular attention to the instructed mechanism.

Compromised Suppression-Induced Forgetting in Mental Disorders Associated With Intrusive Thoughts

The meta-analysis of the nonclinical samples indicated that SIF is a replicable phenomenon in the general population. We had further hypothesized that it may constitute a beneficial coping mechanism to deal with unwanted thoughts and memories. If this were the case, we would expect individuals who find it more difficult to contain intrusive thoughts in their everyday life to be worse at suppression. To test this account, we meta-analyzed groups of participants with mental disorders characterized by intrusive thoughts, or who were subclinical yet potentially susceptible to such issues as indicated by related trait measures.

In line with our hypothesis, this analysis did not provide evidence for a SIF effect across the clinical and subclinical samples, with the exception of repressive copers. Critically, SIF in the (sub)clinical groups was moreover significantly smaller among the group of emotional disorders (i.e., depression and anxiety) than in the healthy control samples.

The moderator analysis also provided no evidence for SIF in (sub)clinical participants following either *direct retrieval suppression* or *thought substitution*. These results may suggest that individuals with emotional disorders do not fundamentally profit from instructed mechanisms. However, only a few studies had employed *thought substitution*, and all of these examined individuals with depression. Thus, results concerning *thought substitution* should be interpreted carefully.

This is particularly the case given the observed variance across studies, suggesting that the effectiveness of this mechanism may hinge on specific features related to the experimental design. For example, in general, thought substitution may be more effective in inducing forgetting, when individuals are provided with substitutes that are unrelated to the unwanted memories (Benoit & Anderson, 2012; Hertel & Calcaterra, 2005) than when they first have to generate the substitutes themselves (Bergström et al., 2009). Retrieving a substitute (e.g., SHIRT) that is strongly related to the unwanted memory (e.g., VEST) may have the unintended consequence of leading to a coactivation and subsequent strengthening of the latter's representation (see also Anderson, Green, & McCulloch, 2000; Bäuml & Hartinger, 2002; Goodmon & Anderson, 2011). To aid depressed individuals with deficient cognitive control, it would thus be critical to carefully consider the nature of the provided substitutes and their relationship to the unwanted memories (Hertel & Calcaterra, 2005).

We observed a large impairment in SIF for participants with PTSD, GAD, or elevated anxiety. These data thus corroborate prior evidence from individual studies that had reported a negative association between SIF and trait anxiety (Benoit et al., 2016; Waldhauser et al., 2018), poor thought control ability (Catarino et al., 2015), depressed mood (Zhang et al., 2016), or rumination (Fawcett et al., 2015). Moreover, a similar pattern has been reported on indirect measures of memory performance (Hertel et al., 2018) and implicit tests (Mary et al., 2020). The pattern is also consistent with several studies that similarly associated clinical phenomena with related deficient control processes at retrieval (e.g., GAD, Kircanski et al., 2016; clinical depression, Groome & Sterkaj, 2010; substance-related and addictive disorders, Stramaccia et al., 2017; eating disorders, Stramaccia et al., 2018). More generally, these results are in line with a recent meta-analysis that associated broader cognitive control deficits with negative thinking (Zetsche, Bürkner, & Schulze, 2018).

However, we found that repressive copers showed stronger SIF than the control samples. This may be consistent with the observation that less anxious individuals are better at intentionally inducing forgetting, given that a repressive coping style has been associated with low trait anxiety (Kim et al., 2007; Myers, 2010). It may also be consistent with the observation that individuals who spontaneously engage in suppressive negative thoughts over the course of their lives display higher SIF (Hulbert & Anderson, 2018). However, results for this *cluster* were based on three studies only, and therefore especially require further corroboration.

The present meta-analyses focused on memory control at the stage of *retrieval*. We consider this a relevant stage for the treatment of psychological disorders, seeing that patients with PTSD, for example, tend to seek help after intrusive memories would already have had time to consolidate (Marks et al., 2018). Due to the delay between initial experience and subsequent treatment, it may often not be feasible to administer potential interventions at earlier stages (see Visser et al., 2018), such as the recently proposed computer game-based treatment (Iyadurai, Blackwell, et al., 2018) or pharmacological interventions (e.g., Zohar, Sonnino, Juven-Wetzler, & Cohen, 2009) that are effective in the early hours prior to the consolidation of traumatic experiences. Similarly, psychological conditions such as depression and anxiety are also characterized by a problematic focus on past memories in the form of rumination (Koval, Kuppens, Allen, & Sheeber, 2012; Michael, Halligan, Clark, & Ehlers, 2007).

The observed relationship between suppression and mental well-being may constitute a conundrum when also considering clinical evidence that ties suppression to negative outcomes in trauma-related disorders (Holmes et al., 2007). We certainly want to emphasize that we do not claim that it is always beneficial to try to suppress unwanted memories. In general, it is critical to engage with negative life experiences and emotions and to integrate them into who we are (see also Biglan, Hayes, & Pistorello, 2008; for examples of negative consequences of suppression in different contexts, see Dalgleish & Yiend, 2006; Le & Impett, 2016; Srivastava, Tamir, McGonigal, John, & Gross, 2009). However, we also think that there are several factors that can reconcile a beneficial take on suppression with the apparently contradicting clinical experience.

First, if the ability to suppress memories is deficient in people who experience intrusive memories, then asking them to suppress an unwanted memory may in fact have the paradoxical effect of aggravating symptoms. That is, for them, attempts to suppress may be bound to fail and thus counterproductively induce rehearsal—and thus strengthening—of unwanted memories.

Second, there seems to be a difference between the notion of suppression as framed in the *think/no-think* literature versus the literature on cognitive–behavioral therapy and emotion regulation (Engen & Anderson, 2018). Specifically, direct retrieval suppression should not be confused with expressive suppression. Expressive suppression refers to the act of voluntarily inhibiting overt expressions of one's emotional states—for instance, facial expressions or tone of voice (Suchy, 2015). This, in turn, has been extensively associated with poorer well-being (Haga, Kraft, & Corby, 2009; Moore, Zoellner, & Mollenholt, 2008).

Finally, as pointed out by Visser, Lau-Zhu, Henson, and Holmes (2018), it may well turn out to be critical what aspects of a memory are being targeted. In particular, these authors argue that it may be beneficial to spare the voluntary access to some declarative components of a trauma memory (e.g., what happened) while attenuating the intrusiveness of unwanted fragments and their aversive emotional component.

These issues notwithstanding, we propose that memory suppression can serve as a mechanism that helps us control the intrusive retrieval of unwanted memories (Anderson et al., 2004; Benoit, Hulbert, Huddleston, & Anderson, 2015; Depue et al., 2007).

Caveats

Importantly, due to the designs of the primary studies, we are not able to infer a causal relationship between reduced SIF and psychopathology. Prospective studies are needed to disentangle whether SIF impairments precede or follow psychopathology, and to determine their potential role as a disorder-maintaining factor. In this respect, the objective difficulty of obtaining large clinical samples and adequate matched controls, combined with the relatively small effect size (at least when studies do not prescribe a specific suppression mechanism), call for joint efforts to investigate such causal relationships.

We also want to point out some limitations of the extant literature and the current meta-analyses. Based on the meta-analytical effect sizes and the large uncertainty associated with them, we note that the primary studies had used fairly low sample sizes overall (on average, about 24 participants per group), and that they thus were certainly low-powered. Note, however, that studies prescribing either thought substitution or direct retrieval suppression yielded considerably greater effect sizes. Furthermore, the limited number of effects in each clinical *cluster* did not allow us to conduct more fine-grained analyses such as for a possible gradient of impairment associated with the severity of the disorders.

Our conclusions should thus be evaluated with respect to the low power of some of the primary data and the relatively high heterogeneity of some of the meta-analytic models. In addition, we had limited means to assess the impact of some of the chosen moderators. For this reason, in future meta-analytic endeavors, it would be worthwhile to examine other aspects of the *think/no-think* task that may constitute boundary conditions to the efficacy of SIF (such as the number of suppressed target memories, specific features of the learning procedure such as the encoding task, the relatedness of cues and targets, and the maximum allotted time for memory retrieval on the final test). A further synthesis of the published literature can also confirm the magnitude of SIF in the broader population (rather than in healthy vs. (sub)clinical populations; e.g., https://osf.io/hmctu).

Finally, methods such as Egger's regression may underestimate the presence of publication bias, especially so in the context of highly heterogeneous models (Renkewitz & Keiner, 2019). (See also Appendix B for the surprising observation that the inclusion of unpublished studies that yielded null results or even reversed effects can lead to a stronger impression of publication bias as estimated by the trim-and-fill procedure.) Therefore, any meta-analytical efforts should ideally be complemented by preregistered, large-scale replication attempts. Nonetheless, we suggest that meta-analyses currently provide the best available evidence on SIF and its disturbance in clinical disorders.

Conclusions

In light of the present results, we suggest that SIF is the replicable hallmark of a process that allows us to voluntarily prevent memory retrieval. Importantly, the observation that this ability is associated with psychological well-being indicates that it may indeed constitute an adaptive coping mechanism. We certainly neither propose that preventing retrieval is always beneficial nor do we suggest that fostering suppression would necessarily be an adequate therapeutic intervention. Yet, in our everyday life, it may help us control intrusive and unwanted thoughts and thus allow us to edit the contents of our memories.

Context

Our lab examines intentional forgetting as an adaptive memory process and has contributed to the understanding of the underlying neural mechanisms (e.g., Benoit & Anderson, 2012; Benoit et al., 2015). Going forward, we think it is important to better characterize the contribution of this process to maintaining mental well-being. As such, we thought it prudent to analyze the literature for an estimate of the effect size of SIF in healthy and (sub)clinical populations.

References

References marked with an asterisk indicate articles included in the analyses.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike. Springer series in statistics (Perspectives in statistics)* (pp. 199–213). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4612-1694-0_15

Algarabel, S., Luciano, J. V., & Martínez, J. L. (2006). Inhibitory voluntary control of memory: effect of stimulus onset asynchrony on reaction time to suppressed memories. *International Journal of Methodology and Experimental Psychology, 27,* 57–77.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

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

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1141–1159. http://dx.doi.org/10.1037/0278-7393.26.5.1141

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

Anderson, M. C., & Huddleston, E. (2012). Towards a cognitive and neurobiological model of motivated forgetting. In *True and false recovered memories* (pp. 53–120). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-1195-6_3

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. http://dx.doi.org/10.1126/science.1089504

Bäuml, K. H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10,* 215–224.

Belli, R. F. (Ed.). (2011). *True and false recovered memories: Toward a reconciliation of the debate* (Vol. 58). New York, NY: Springer Science & Business Media.

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

Benoit, R. G., Davies, D. J., & Anderson, M. C. (2016). Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences of the United States of America, 113,* E8492–E8501. http://dx.doi.org/10.1073/pnas.1606604114

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. http://dx.doi.org/10.1162/jocn_a_00696

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. http://dx.doi.org/10.1016/j.neuroimage.2009.06.051

Bergström, Z. M., Velmans, M., de Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. *Brain Research, 1151,* 119–133. http://dx.doi.org/10.1016/j.brainres.2007.03.014

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174,* 727–735. http://dx.doi.org/10.1097/00005053-198612000-00004

Biglan, A., Hayes, S. C., & Pistorello, J. (2008). Acceptance and commitment: Implications for prevention science. *Prevention Science, 9,* 139–152. http://dx.doi.org/10.1007/s11121-008-0099-4

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5,* 176–196. http://dx.doi.org/10.1006/ccog.1996.0011

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9,* 255–268. http://dx.doi.org/10.1016/S0022-5371(70)80059-7

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: Foundations for a theory of posttraumatic stress disorder. *Psychological Bulletin, 140,* 69–97. http://dx.doi.org/10.1037/a0033722

Briggs, G. E. (1954). Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology, 47,* 285–293. http://dx.doi.org/10.1037/h0060251

Bulevich, J. B., Roediger, H. L., III, Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & Cognition, 34,* 1569–1577. http://dx.doi.org/10.3758/BF03195920

Card, N. A. (2015). *Applied meta-analysis for social science research*. New York, NY: Guilford Press Publications.

*Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychological Science, 26,* 604–616. http://dx.doi.org/10.1177/0956797615569889

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101–129. http://dx.doi.org/10.2307/3001666

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman and Hall.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. New York, NY: Russel Sage Foundation. http://dx.doi.org/10.7758/9781610448864

Dalgleish, T., & Yiend, J. (2006). The effects of suppressing a negative autobiographical memory on concurrent intrusions and subsequent autobiographical recall in dysphoria. *Journal of Abnormal Psychology, 115,* 467–473. http://dx.doi.org/10.1037/0021-843X.115.3.467

*Depue, B. E., Burgess, G. C., Willcutt, E. G., Ruzic, L., & Banich, M. T. (2010). Inhibitory control of memory retrieval and motor processing associated with the right lateral prefrontal cortex: Evidence from deficits in individuals with ADHD. *Neuropsychologia, 48,* 3909–3917. http://dx.doi.org/10.1016/j.neuropsychologia.2010.09.013

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science, 317,* 215–219. http://dx.doi.org/10.1126/science.1139560

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia, 51,* 2371–2388. http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.017

*Dieler, A. C., Herrmann, M. J., & Fallgatter, A. J. (2014). Voluntary suppression of thoughts is influenced by anxious and ruminative tendencies in healthy volunteers. *Memory, 22,* 184–193. http://dx.doi.org/10.1080/09658211.2013.774420

*Diwadkar, V. A., Re, M., Cecchetto, F., Garzitto, M., Piccin, S., Bonivento, C., . . . Brambilla, P. (2017). Attempts at memory control induce dysfunctional brain activation profiles in generalized anxiety disorder: An exploratory fMRI study. *Psychiatry Research: Neuroimaging, 266,* 42–52. http://dx.doi.org/10.1016/j.pscychresns.2017.05.010

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification, 41,* 323–339. http://dx.doi.org/10.1177/0145445516673998

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

Ehlers, A., Hackmann, A., & Michael, T. (2004). Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory, 12,* 403–415. http://dx.doi.org/10.1080/09658210444000025

Engen, H. G., & Anderson, M. C. (2018). Memory control: A fundamental mechanism of emotion regulation. *Trends in Cognitive Sciences, 22,* 982–995. http://dx.doi.org/10.1016/j.tics.2018.07.015

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6,* 21–29. http://dx.doi.org/10.1016/0191-8869(85)90026-1

*Fawcett, J. M., Benoit, R. G., Gagnepain, P., Salman, A., Bartholdy, S., Bradley, C., . . . Anderson, M. C. (2015). The origins of repetitive thought in rumination: Separating cognitive style from deficits in inhibitory control over memory. *Journal of Behavior Therapy and Experimental Psychiatry, 47,* 1–8. http://dx.doi.org/10.1016/j.jbtep.2014.10.009

Fawcett, J. M., & Hulbert, J. C. (2020). The many faces of forgetting: Toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory & Cognition, 9,* 1–18. http://dx.doi.org/10.1016/j.jarmac.2019.11.002

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review, 34,* 402–416. http://dx.doi.org/10.1016/j.cpr.2014.06.001

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 of the United States of America, 111*(13), E1310–E1319.

Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal*

of Experimental Psychology: Learning, Memory, and Cognition, 37, 416–436. http://dx.doi.org/10.1037/a0021963

Goodwin, G. M. (2015). The overlap between anxiety, depression, and obsessive-compulsive disorder. *Dialogues in Clinical Neuroscience, 17,* 249–260.

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: Advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research, 23,* 41–57. http://dx.doi.org/10.1002/mpr.1410

Groome, D., & Sterkaj, F. (2010). Retrieval-induced forgetting and clinical depression. *Cognition and Emotion, 24,* 63–70. http://dx.doi.org/10.1080/02699930802536219

Hackmann, A., Ehlers, A., Speckens, A., & Clark, D. M. (2004). Characteristics and content of intrusive memories in PTSD and their changes with treatment. *Journal of Traumatic Stress, 17,* 231–240. http://dx.doi.org/10.1023/B:JOTS.0000029266.88369.fd

Haga, S. M., Kraft, P., & Corby, E. K. (2009). Emotion regulation: Antecedents and well-being outcomes of cognitive reappraisal and expressive suppression in cross-cultural samples. *Journal of Happiness Studies, 10,* 271–291. http://dx.doi.org/10.1007/s10902-007-9080-3

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis: Vol. 5. No. 3* (pp. 207–219). Upper Saddle River, NJ: Prentice hall.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. http://dx.doi.org/10.1002/jrsm.5

Hertel, P. T. (1997). On the contributions of deficient cognitive control to memory impairments in depression. *Cognition and Emotion, 11,* 569–583. http://dx.doi.org/10.1080/026999397379890a

Hertel, P. T. (1998). Relation between rumination and impaired memory in dysphoric moods. *Journal of Abnormal Psychology, 107,* 166–172. http://dx.doi.org/10.1037/0021-843X.107.1.166

Hertel, P. T. (2007). Impairments in inhibition or cognitive control in psychological disorders. *Applied & Preventive Psychology, 12,* 149–153. http://dx.doi.org/10.1016/j.appsy.2007.09.006

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review, 12,* 484–489. http://dx.doi.org/10.3758/BF03193792

*Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological Science, 14,* 573–578. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1467.x

Hertel, P. T., Large, D., Stück, E. D., & Levy, A. (2012). Suppression-induced forgetting on a free-association test. *Memory, 20,* 100–109. http://dx.doi.org/10.1080/09658211.2011.647036

*Hertel, P. T., & Mahan, A. (2008). Depression-related differences in learning and forgetting responses to unrelated cues. *Acta Psychologica, 127,* 636–644. http://dx.doi.org/10.1016/j.actpsy.2007.11.004

Hertel, P. T., Maydon, A., Ogilvie, A., & Mor, N. (2018). Ruminators (unlike others) fail to show suppression-induced forgetting on indirect measures of memory. *Clinical Psychological Science, 6,* 872–881. http://dx.doi.org/10.1177/2167702618785339

*Hertel, P., & McDaniel, L. (2010). The suppressive power of positive thinking: Aiding suppression-induced forgetting in repressive coping. *Cognition and Emotion, 24,* 1239–1249. http://dx.doi.org/10.1080/02699930903172377

Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science, 11,* 359–372. http://dx.doi.org/10.1177/1745691616635594

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions*. New York, NY: Wiley.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557

Holmes, E. A., Moulds, M. L., Kavanagh, D., Depue, B. E., Curran, T., & Banich, M. T. (2007). Memory suppression in PTSD treatment? *Science, 318,* 1722. http://dx.doi.org/10.1126/science.318.5857.1722a

Hulbert, J. C., & Anderson, M. C. (2018). What doesn't kill you makes you stronger: Psychological trauma and its relationship to enhanced memory control. *Journal of Experimental Psychology: General, 147,* 1931–1949. http://dx.doi.org/10.1037/xge0000461

Hulbert, J. C., & Anderson, M. C. (2020). Does retrieving a memory insulate it against memory inhibition? A retroactive interference study. *Memory, 28,* 293–308. http://dx.doi.org/10.1080/09658211.2019.1710216

Hulbert, J. C., Hirschstein, Z., Brontë, C. A. L., & Broughton, E. (2018). Unintended side effects of a spotless mind: Theory and practice. *Memory, 26,* 306–320. http://dx.doi.org/10.1080/09658211.2017.1354999

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *British Medical Journal Open, 6,* e010247. http://dx.doi.org/10.1136/bmjopen-2015-010247

Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., . . . Holmes, E. A. (2018). Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: A proof-of-concept randomized controlled trial. *Molecular Psychiatry, 23,* 674–682. http://dx.doi.org/10.1038/mp.2017.23

Iyadurai, L., Visser, R. M., Lau-Zhu, A., Porcheret, K., Horsch, A., Holmes, E. A., & James, E. L. (2018). Intrusive memories of trauma: A target for research bridging cognitive science and its clinical application. *Clinical Psychology Review, 69,* 67–82.

*Joormann, J., Hertel, P. T., Brozovich, F., & Gotlib, I. H. (2005). Remembering the good, forgetting the bad: Intentional forgetting of emotional material in depression. *Journal of Abnormal Psychology, 114,* 640–648. http://dx.doi.org/10.1037/0021-843X.114.4.640

*Joormann, J., Hertel, P. T., LeMoult, J., & Gotlib, I. H. (2009). Training forgetting of negative material in depression. *Journal of Abnormal Psychology, 118,* 34–43. http://dx.doi.org/10.1037/a0013794

*Kim, D. Y., Oh, D. H., Kim, S. H., Sim, K. B., & Lee, J. H. (2013). Effects of intentional suppression of recall of unwanted images in repressors and nonrepressors. *Social Behavior and Personality, 41,* 319–326. http://dx.doi.org/10.2224/sbp.2013.41.2.319

*Kim, K., Yi, D., Yang, E., & Lee, K. (2007). What makes repressors good suppressors? The effect of trait anxiety. *Korean Journal of Psychology, 26,* 261–277.

*Kircanski, K., Johnson, D. C., Mateen, M., Bjork, R. A., & Gotlib, I. H. (2016). Impaired retrieval inhibition of threat material in generalized anxiety disorder. *Clinical Psychological Science, 4,* 320–327. http://dx.doi.org/10.1177/2167702615590996

Kircanski, K., Joormann, J., & Gotlib, I. H. (2012). Cognitive aspects of depression. *Wiley Interdisciplinary Reviews: Cognitive Science, 3,* 301–313. http://dx.doi.org/10.1002/wcs.1177

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2,* 61–76. http://dx.doi.org/10.1002/jrsm.35

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion, 26,* 1412–1427. http://dx.doi.org/10.1080/02699931.2012.667392

*Küpper, C. S., Benoit, R. G., Dalgleish, T., & Anderson, M. C. (2014). Direct suppression as a mechanism for controlling unpleasant memories in daily life. *Journal of Experimental Psychology: General, 143,* 1443–1449. http://dx.doi.org/10.1037/a0036518

Le, B. M., & Impett, E. A. (2016). The costs of suppressing negative emotions and amplifying positive emotions during parental caregiving. *Personality and Social Psychology Bulletin, 42,* 323–336. http://dx.doi.org/10.1177/0146167216629122

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Luciano, J. V., Algarabel, S., Tomás, J. M., & Martínez, J. L. (2005). Development and validation of the thought control ability questionnaire. *Personality and Individual Differences, 38,* 997–1008. http://dx.doi.org/10.1016/j.paid.2004.06.020

Marks, E. H., Franklin, A. R., & Zoellner, L. A. (2018). Can't get it out of my mind: A systematic review of predictors of intrusive memories of distressing events. *Psychological Bulletin, 144,* 584–640. http://dx.doi.org/10.1037/bul0000132

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., . . . Gagnepain, P. (2020). Resilience after trauma: The role of memory suppression. *Science, 367,* eaay8477. http://dx.doi.org/10.1126/science.aay8477

*Marzi, T., Regina, A., & Righi, S. (2014). Emotions shape memory suppression in trait anxiety. *Frontiers in Psychology, 4,* 1001. http://dx.doi.org/10.3389/fpsyg.2013.01001

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12,* 227–255. http://dx.doi.org/10.1016/0272-7358(92)90116-P

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39,* 352–370. http://dx.doi.org/10.1037/h0069819

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46. http://dx.doi.org/10.1037/1082-989X.1.1.30

McTeague, L. M., Goodkind, M. S., & Etkin, A. (2016). Transdiagnostic impairment of cognitive control in mental illness. *Journal of Psychiatric Research, 83,* 37–46. http://dx.doi.org/10.1016/j.jpsychires.2016.08.001

Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain Research, 1255,* 132–147. http://dx.doi.org/10.1016/j.brainres.2008.11.073

Michael, T., Halligan, S. L., Clark, D. M., & Ehlers, A. (2007). Rumination in posttraumatic stress disorder. *Depression and Anxiety, 24,* 307–317. http://dx.doi.org/10.1002/da.20228

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151,* 264–269. http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135

Moore, S. A., Zoellner, L. A., & Mollenholt, N. (2008). Are expressive suppression and cognitive reappraisal associated with stress-related symptoms? *Behaviour Research and Therapy, 46,* 993–1000. http://dx.doi.org/10.1016/j.brat.2008.05.001

Myers, L. B. (2010). The importance of the repressive coping style: findings from 30 years of research. *Anxiety, Stress & Coping, 23,* 3–17.

Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26,* 1253–1274. http://dx.doi.org/10.1007/s10682-012-9555-5

*Nemeth, V. L., Kurgyis, E., Csifcsak, G., Maraz, A., Almasi, D. A., Drotos, G., . . . Must, A. (2014). The impact of intermediate-term alcohol abstinence on memory retrieval and suppression. *Frontiers in Psychology, 5,* 1396. http://dx.doi.org/10.3389/fpsyg.2014.01396

Nørby, S. (2015). Why forget? On the adaptive value of memory loss. *Perspectives on Psychological Science, 10,* 551–578. http://dx.doi.org/10.1177/1745691615596787

Nørby, S. (2018). Forgetting and emotion regulation in mental health, anxiety and depression. *Memory, 26,* 342–363. http://dx.doi.org/10.1080/09658211.2017.1346130

*Noreen, S., Cooke, R., & Ridout, N. (2019). Investigating the mediating effect of working memory on intentional forgetting in dysphoria. *Psychological Research*. Advance online publication. http://dx.doi.org/10.1007/s00426-019-01225-y

*Noreen, S., & Ridout, N. (2016a). Examining the impact of thought substitution on intentional forgetting in induced and naturally occurring dysphoria. *Psychiatry Research, 241,* 280–288. http://dx.doi.org/10.1016/j.psychres.2016.04.086

*Noreen, S., & Ridout, N. (2016b). Intentional forgetting in dysphoria: Investigating the inhibitory effects of thought substitution using independent cues. *Journal of Behavior Therapy and Experimental Psychiatry, 52,* 110–118. http://dx.doi.org/10.1016/j.jbtep.2016.04.002

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review, 114,* 887.

Payne, L., & Sekuler, R. (2014). The importance of ignoring: Alpha oscillations protect selectivity. *Current Directions in Psychological Science, 23,* 171–177. http://dx.doi.org/10.1177/0963721414529145

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association, 295,* 676–680. http://dx.doi.org/10.1001/jama.295.6.676

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology, 61,* 991–996. http://dx.doi.org/10.1016/j.jclinepi.2007.11.010

Racsmány, M., Conway, M. A., Keresztes, A., & Krajcsi, A. (2012). Inhibition and interference in the think/no-think task. *Memory & Cognition, 40,* 168–176. http://dx.doi.org/10.3758/s13421-011-0144-6

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research: A comparative evaluation of six statistical methods. *Zeitschrift für Psychologie, 227,* 261–279. http://dx.doi.org/10.1027/2151-2604/a000386

Rohatgi, A. (2017). WebPlotDigitizer (Version 3.11) [Computer software]. Retrieved from http://arohatgi.info/WebPlotDigitizer

*Ryckman, N. (2015). Exploring intentional forgetting: Using novel think/no-think paradigms to investigate memory suppression (Doctoral dissertation). School of Psychology, University of Auckland, Auckland, New Zealand.

*Sacchet, M. D., Levy, B. J., Hamilton, J. P., Maksimovskiy, A., Hertel, P. T., Joormann, J., . . . Gotlib, I. H. (2017). Cognitive and neural consequences of memory suppression in major depressive disorder. *Cognitive, Affective & Behavioral Neuroscience, 17,* 77–93. http://dx.doi.org/10.3758/s13415-016-0464-x

Sahakyan, L., Waldum, E. R., Benjamin, A. S., & Bickett, S. P. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory & Cognition, 37,* 464–476. http://dx.doi.org/10.3758/MC.37.4.464

*Salamé, P., & Danion, J. M. (2007). Inhibition of inappropriate responses is preserved in the think-no-think and impaired in the random number generation tasks in schizophrenia. *Journal of the International Neuropsychological Society, 13,* 277–287. http://dx.doi.org/10.1017/S1355617707070300

Spiess, A-N. (2018). qpcR: Modelling and analysis of real-time PCR data (R package version 1.4 -1) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/qpcR/index.html

Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: A prospective study of the transition to college. *Journal of Personality and Social Psychology, 96,* 883–897. http://dx.doi.org/10.1037/a0014755

*Stephens, E., Braid, A., & Hertel, P. T. (2013). Suppression-induced reduction in the specificity of autobiographical memories. *Clinical Psy-*

chological Science, 1, 163–169. http://dx.doi.org/10.1177/2167702612467773

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. Rothstein, A. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Chichester, England: Wiley. http://dx.doi.org/10.1002/0470870168.ch6

Stramaccia, D. F., Penolazzi, B., Libardi, A., Genovese, A., Castelli, L., Palomba, D., & Galfano, G. (2018). Control over interfering memories in eating disorders. *Journal of Clinical and Experimental Neuropsychology, 40,* 30–44. http://dx.doi.org/10.1080/13803395.2017.1313392

Stramaccia, D. F., Penolazzi, B., Monego, A. L., Manzan, A., Castelli, L., & Galfano, G. (2017). Suppression of competing memories in substance-related and addictive disorders: A retrieval-induced forgetting study. *Clinical Psychological Science, 5,* 410–417. http://dx.doi.org/10.1177/2167702616671780

Suchy, Y. (2015). *Executive functioning: A comprehensive guide for clinical practice*. New York, NY: Oxford University Press.

Sullivan, D. R., Marx, B., Chen, M. S., Depue, B. E., Hayes, S. M., & Hayes, J. P. (2019). Behavioral and neural correlates of memory suppression in PTSD. *Journal of Psychiatric Research, 112,* 30–37. http://dx.doi.org/10.1016/j.jpsychires.2019.02.015

Thorndike, E. L. (1913). *The psychology of learning*. New York, NY: Teachers College, Columbia University.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53,* 207–216. http://dx.doi.org/10.1016/S0895-4356(99)00161-4

Tomlinson, T. D., Huber, D. E., Rieth, C. A., & Davelaar, E. J. (2009). An interference account of cue-independent forgetting in the no-think paradigm. *Proceedings of the National Academy of Sciences, 106,* 15588–15593.

Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist, 62,* 74–82.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60. http://dx.doi.org/10.1037/h0044616

van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory, 25,* 1201–1216. http://dx.doi.org/10.1080/09658211.2017.1282518

Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1433–1448. http://dx.doi.org/10.1037/a0032975

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. Advance online publication. http://dx.doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1,* 112–125. http://dx.doi.org/10.1002/jrsm.11

Visser, R. M., Lau-Zhu, A., Henson, R. N., & Holmes, E. A. (2018). Multiple memory systems, multiple time points: How science can inform treatment to control the expression of unwanted emotional memories. *Philosophical Transactions Royal Society B, 373,* 20170209. http://dx.doi.org/10.1098/rstb.2017.0209

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. http://dx.doi.org/10.3758/BF03206482

*Waldhauser, G. T., Dahl, M. J., Ruf-Leuschner, M., Müller-Bamouh, V., Schauer, M., Axmacher, N., . . . Hanslmayr, S. (2018). The neural dynamics of deficient memory control in heavily traumatized refugees. *Scientific Reports, 8,* 13132. http://dx.doi.org/10.1038/s41598-018-31400-x

Wang, Y., Luppi, A., Fawcett, J., & Anderson, M. C. (2019). Reconsidering unconscious persistence: Suppressing unwanted memories reduces their indirect expression in later thoughts. *Cognition, 187,* 78–94. http://dx.doi.org/10.1016/j.cognition.2019.02.016

Wessel, I., Albers, C. J., Zandstra, A. E., & Heininga, V. E. (2020, March 27). A multiverse analysis of early at-tempts to replicate memory suppression with the think/no-think task. *PsyArXiv*. http://dx.doi.org/10.31234/osf.io/2dr9e

*Wessel, I., Wetzels, S., Jelicic, M., & Merckelbach, H. (2005). Dissociation and memory suppression: A comparison of high and low dissociative individuals' performance on the think–no think task. *Personality and Individual Differences, 39,* 1461–1470. http://dx.doi.org/10.1016/j.paid.2005.05.009

Williams, A. D., Moulds, M. L., Grisham, J. R., Gay, P., Lang, T., Kandris, E., . . . Yap, C. (2010). A psychometric evaluation of the Thought Control Ability Questionnaire (TCAQ) and the prediction of cognitive control. *Journal of Psychopathology and Behavioral Assessment, 32,* 397–405. http://dx.doi.org/10.1007/s10862-009-9171-z

World Health Organization. (2018). *International statistical classification of diseases and related health problems (11th Revision)*. Geneva, Switzerland: Author. Retrieved from https://icd.who.int/browse11/l-m/en

Zetsche, U., Bürkner, P. C., & Schulze, L. (2018). Shedding light on the association between repetitive negative thinking and deficits in cognitive control: A meta-analysis. *Clinical Psychology Review, 63,* 56–65. http://dx.doi.org/10.1016/j.cpr.2018.06.001

*Zhang, D., Xie, H., Liu, Y., & Luo, Y. (2016). Neural correlates underlying impaired memory facilitation and suppression of negative material in depression. *Scientific Reports, 6,* 37556. http://dx.doi.org/10.1038/srep37556

Zohar, J., Sonnino, R., Juven-Wetzler, A., & Cohen, H. (2009). Can posttraumatic stress disorder be prevented. *CNS Spectrums, 14*(1, Suppl. 1), 44–51.

(*Appendices follow*)

Appendix A

Simple Random-Effects Models of Only Independent Effect Sizes

Many of the included studies provided multiple estimates of SIF (i.e., one for each level of within-subject manipulations such as the number of repetitions). The main analysis is based on a multilevel random-effects model that comprehensively includes all of these estimates.

Here, we provide the results of complementary simpler random-effects models that include only one independent effect size from each independent group in each study. In particular, in case of multiple dependent effect sizes, we selected the one that constituted the standard manipulation (e.g., greater number of suppression repetitions) and/or outcome measure (binary rather than continuous measures of recall accuracy) in respect to the extant literature on the *think/no-think* task (as in a prior preprint on PsyArXiv doi: 10.31234/osf.io/5wynm; see Card, 2015, pp. 192–193; Cooper, Hedges, & Valentine, 2019, p. 282; Higgins & Green, 2011, Chapter 3; Lipsey & Wilson, 2001, p. 125). This approach presumably provides an estimate of the strongest manipulations (e.g., a greater rather than a lesser number of suppression repetitions). We perform these models separately for the *healthy* samples and the *(sub)clinical* samples. This approach also allows us to implement the trim-and-fill procedure (Duval & Tweedie, 2000), in a further attempt to gauge the degree of publication bias. This procedure estimates the number of missing studies in the meta-analytic model due to publication bias and the impact that they might have on the meta-analytic effect size.

Suppression-Induced Forgetting in Healthy Samples

For the *healthy* samples (32 effects), there was a significant SIF effect of 0.27, 95% CI [0.16, 0.39], 95% PI [−.22, .77], $p < .001$, with a large, significant heterogeneity, $I^2 = 58.05\%$, $Q_{(31)} = 76.83$, $p < .001$. Following trim and fill, with the insertion of data points for an estimated seven missing studies, the model remained significant with a reduced estimate of 0.17, 95% CI [0.05, 0.30], 95% PI [−.44, .79], $p < .001$, and a large heterogeneity, $I^2 = 66.63\%$ which was also significant, $Q_{(38)} = 107.29$, $p < .001$.

No Suppression-Induced Forgetting in (Sub)Clinical Samples

For the (*sub)clinical* samples (32 effects), there was a nonsignificant SIF effect of 0.08, 95% CI [−0.08, 0.23], 95% PI [−.68, .84], $p = .320$, with a large, significant heterogeneity, $I^2 = 76.16\%$, $Q_{(31)} = 118.27$, $p < .001$. After trim and fill (with the insertion of data points for an estimated one missing study), the model remained nonsignificant with a reduced estimate of 0.06, 95% CI [−0.10, 0.21], 95% PI [−.73, .84], $p < .001$, and a large, significant heterogeneity, $I^2 = 77.40\%$, $Q_{(38)} = 126.53$, $p < .001$.

(*Appendices continue*)

Appendix B

Simple Random-Effects Models of Average Effect Sizes

To perform a trim-and-fill adjustment across all reported effect sizes (rather than the most standard measures of SIF), we first averaged all interdependent effect sizes of each study (i.e., the within-subject levels of a given independent variable). We then fitted separate simple random-effects models based on these average effect sizes.

Suppression-Induced Forgetting in Healthy Samples: Stronger Correction

For the *healthy* samples (32 effects), there was a significant SIF effect of 0.26, 95% CI [0.14, 0.39], 95% PI [−.30, .82], $p < .001$, and large heterogeneity, $I^2 = 64.05\%$ which was also significant, $Q_{(31)} = 85.22$, $p < .001$.

The trim-and-fill correction of the data from all studies added data points for an estimated eight missing studies, which resulted in a trend only for an effect of 0.13, 95% CI [−0.02, 0.27], 95% PI [−.65, .90], $p = .084$, with a large heterogeneity, $I^2 = 75.70\%$ which was also significant, $Q_{(39)} = 133.96$, $p < .001$.

We further sought to gauge the actual publication bias by refitting the trim-and-fill model based only on the published literature. This analysis thus excludes the three experiments (Experiment 2, plus the combined Experiments 3 and 4) by Ryckman (2015) that either yielded a null effect or, in fact, reversed SIF. Somewhat surprisingly, the trim-and-fill procedure exerted a lesser penalty on the estimate for only the published studies—with the insertion of six additional data points. Indeed, SIF remained significant with an effect of 0.18, 95% CI [0.03, 0,32], 95% PI [−0.55, 0.90], $p = .016$, with a large and significant heterogeneity, $Q_{(35)} = 107.03$, $p < .001$.

No Suppression-Induced Forgetting in (Sub)Clinical Samples

For the *clinical* samples (32 effects), SIF was not significant with an effect of 0.03 (95% CI [−0.11, 0.17], 95% PI [−.64, .69], $p = .719$, and large, significant heterogeneity, $I^2 = 71.48\%$, $Q_{(31)} = 85.22$, $p < .001$).

The trim-and-fill correction of the averaged effect sizes did not yield any differences for the *clinical* samples.

In keeping with the same reasoning as for the *healthy* samples, we then removed the unpublished studies by Ryckman (2015) and reestimated the trim-and-filled model. The procedure did not add any data points to the funnel plot and SIF was also not significant with an effect of 0.03, 95% CI [−0.12, 0.18], 95% PI [−.67, .74], $p = .688$, and large, significant heterogeneity, $I^2 = 71.52\%$, $Q_{(29)} = 99.52$, $p < .001$.

Received May 27, 2019
Revision received June 10, 2020
Accepted July 16, 2020 ■