Forgetting as a Consequence of Retrieval Suppression:

A Meta-Analytic Review of the Think/No-Think Paradigm

by © Chris Clark

A Thesis submitted to the School of Graduate Studies in partial

fulfilment of the requirement for the degree of

Master of Science in Experimental Psychology

Department of Psychology

Memorial University of Newfoundland

2021

St. John’s, Newfoundland and Labrador

Abstract

The present thesis used meta-analytic methods to define the typical magnitude of the suppression induced forgetting (SIF) effect produced by the Think/No-Think paradigm as well as assess several moderators. A literature search was conducted using seven online databases, review papers and discussion with experts in the field to identify 82 relevant studies. Using Bayesian multi-level modelling techniques to provide a synthesis of the existing literature it was determined that the typical Think/No-Think study produced a small SIF effect. This means that the Think/No-Think paradigm decreases recall accuracy for no-think items relative to baseline items, with broad heterogeneity in the observed effect across studies depending on the methods. Additionally, it was determined that the SIF effect was larger (a) in studies using same probe (SP) tests than independent probe (IP) tests; (b) when the data were conditionalized on successful learning of the stimuli pairs during the test-feedback phase; (c) when participants were given more specific instructions on how to avoid thinking of the no-think items, and; (d) at higher numbers of repetitions (for unconditionalized data). Neither stimulus emotional content nor stimuli type impacted the magnitude of SIF. These findings provide strong evidence of the viability of the SIF effect while also informing future theoretical discussions concerning its mechanisms. Several recommendations are made for future Think/No-Think studies concerning sample size and methodological decisions in an effort to improve the precision and reliability of findings.

General Summary

The aim of this thesis was to determine if the task known as Think/No-Think (TNT) is capable of consistently producing an effect on memory, as well as identifying specific factors that may make the effect smaller or larger. This task asks participants to memorize pairs of stimuli followed by asking them to either think of, or avoid thinking of, one half of the pair when given the other half. Participants are then asked to remember the pairs, with the expected result being that items they were told to avoid thinking about being harder to remember than those they were allowed to think about. This is known as the suppression induced forgetting (SIF) effect. To this end, a review of published and unpublished research including the TNT task was conducted to determine its average effect on memory. It was found that the average TNT study shows a reliable SIF effect, meaning participants were able to forget specific memories when directed. Additionally, it was found that the effect was larger when participants (a) were tested by giving them one of the items directly and asking for its pair; (b) were given more specific instructions on how to avoid thinking about the items, and; (c) had a greater number of opportunities to avoid thinking about the word. Overall, this thesis suggests that people are capable of intentionally forgetting specific memories within the TNT task.

Acknowledgements

I would like to first thank my supervisor, Dr. Jonathan Fawcett, for continuing to be an excellent teacher and mentor throughout an unexpected and unprecedented time. I thoroughly enjoyed my time working with Jon and look forward to applying the numerous skills and experiences I have gained as his student wherever the future takes me. I would also like to thank my parents, Dana and Denise Clark, for a lifetime of support and always driving me to accomplish my goals. Finally, I would like to thank my family and friends both here on the island and across the country for their support and help in keeping me (relatively) sane through the completion of this project and my degree as a whole.

Table of Contents

List of Tables

List of Figures

Forgetting as a Consequence of Retrieval Suppression: A Meta-Analytic Review of the Think/No-Think Paradigm

Memory is one of the most crucial aspects of cognition, playing a role in subtle underlying processes and everyday cognitive functioning such as remembering dates, times, names, and general life events. As such, the study of memory has long played a significant role in the study of cognition with some of the most well researched memory effects involving improving and maintaining an individual's memory capabilities (e.g., Craik & Watkins, 1973; Levy, 1996; MacLeod et al., 2010; Verhaeghen et al., 1992). However, a somewhat less studied, yet equally important, area of cognitive research is the study of forgetting, or how information is removed from memory. The act of forgetting has historically been portrayed as an autonomous, passive process unaffected by conscious intention (e.g., Brown, 1958). Contrary to this belief, theorists in the realms of cognition and cognitive neuroscience have long held that there exist direct neurocognitive mechanisms that act to selectively influence the strength of specific memories, limiting successful recall.

The origins of these theories can be traced as far back as the original two factor theory of interference that implied that memories could be unlearned to prevent future intrusions (Melton & Irwin, 1940). Similarly, Freud argued that memories could be selectively repressed to reduce psychological conflicts (Breuer & Freud, 1955). More modern proponents of these theories argue that forgetting, particularly when intentional, contributes broadly to the maintenance and efficiency of cognition (e.g., Anderson & Hulbert, 2021; Fawcett & Hulbert, 2020; Norby, 2018). However, there remains some debate concerning the exact mechanisms employed during intentional forgetting, as well as the exact efficacy of intentional forgetting processes. To provide clarity to this discussion, the present thesis seeks to examine the ability to intentionally suppress retrieval of specific memories in the think/no think paradigm as well as to examine several key moderators that may alter the effectiveness of retrieval suppression.

Intentional Forgetting

Intentional forgetting can be broadly defined as a collection of cognitive processes which are engaged by the conscious and directed desire to forget previously encoded information. A review of the cognitive neuroscience literature surrounding intentional forgetting conducted by Anderson and Hulbert (2021) identified that these processes include the disruption of memory encoding and retrieval. A variety of laboratory paradigms have been utilized to empirically study these processes, including but not limited to item–method and list-method directed forgetting (MacLeod, 1999), retrieval-induced forgetting (Anderson et al., 1994) and the think/no think paradigm (Anderson & Green, 2001). Each of these paradigms examines distinct aspects of intentional forgetting, thus it is important that they be described in sufficient detail to differentiate their contributions to the literature. As such, I will first provide a brief overview of the directed forgetting and retrieval-induced forgetting paradigms before providing a more in-depth description of the think/no-think paradigm as the primary focus of the present thesis.

Directed Forgetting

The directed forgetting paradigm was first formally defined in the early 1970's (Bjork & Woodward, 1973) and involves participants being presented with a series of items, typically words (e.g., DePrince & Freyd, 2004) or pictures (e.g., Quinlan et al., 2010), followed by an instruction to forget or remember specific items. Next, participants complete a test of free recall (i.e., asking participants to name or describe as many stimuli as they can; e.g., Bailey & Chapman, 2012), or recognition (i.e., asking participants to classify stimuli as new or old; e.g., Sahakyan et al., 2009). Originally, the directed forgetting paradigm consisted of a unitary effect which was largely thought to be caused by selective rehearsal (Bjork, 1972). Over time the theory behind the effect dissociated into two main areas of focus, resulting in two distinct versions of the directed forgetting paradigm depending on the timing with which the participants receive the remember and forget instructions (Basden et al., 1993; MacLeod, 1999).

In list-method directed forgetting participants are typically presented with an entire list of words before being told to remember or forget the list as a whole (e.g., Zellner & Bauml, 2006). This method is frequently utilized in between-subjects designs wherein participants are presented with two lists of words and instructed to forget or remember the first list depending on the assigned condition. List-method directed forgetting consistently results in decreased recall of forget lists relative to remember lists (Macleod, 1999). Contrary to this finding, list-method directed forgetting commonly fails to produce a difference in recognition between remember and forget items (Golding & MacLeod, 1998). However, some theorists hold that these studies focus exclusively on recognition of items presented pre-cue, demonstrating that there is a directed forgetting effect for post-cue items (Benjamin, 2006). Specifically, items following a cue to forget the preceding items show increased recognition accuracy, potentially indicating a benefit to remember items as opposed to reduced memory of forget items. Alternatively, some theorists argue that the inability to observe an effect on recognition is due to methodological issues in testing as opposed to an inherent issue with list-method directed forgetting (Sahakyan et al., 2009).

List-method directed forgetting has also been examined using modified versions of the typical paradigm described above. Sahakyan et al. (2013) summarized that the typical method confounds the memory instruction with other factors such as proactive interference between the first and second list, wherein there may be interference for the second list but not the first. As a result, multiple alternative methods have been utilized to mitigate these complications including the three-list design (Lehman & Malmberg, 2009) wherein participants are told to remember or forget the second list as opposed to the first. This theoretically allows for proactive interference to equally influence both the 'target' (2^{nd}) and 'control' (3^{rd}) lists. Alternatively, the four-lists design proposed by Zellner and Bauml (2006) utilizes a within-subjects manipulation by using two-blocks of two list testing as well as counterbalanced order of testing to reduce order effects. However, there are some concerns that this method allows for the refinement of encoding strategies in the second block of lists, introducing an additional confound (Delaney & Knowles, 2005). While the viability of the various different methods are discussed throughout the literature, it should be noted that list-method directed forgetting consistently produces the same effect regardless of the method employed.

As mentioned previously, one of the earliest theories behind directed forgetting was the selective rehearsal account which posited that the effect was the result of participants preferentially rehearsing the remember items at the cost of the forget items (Bjork, 1972). While this explanation may be sufficient to explain the difference in the recall accuracy between the two lists it fails to provide an answer for the lack of difference in recognition for list one in list-method directed forgetting which can be commonly observed. Later theories began to focus primarily on inhibitory mechanisms that selectively reduce the success of the encoding or retrieval of stimuli in the forget list (Bjork, 1989). However, some researchers have argued that this inhibitory perspective is unable to identify what is specifically inhibited during the presentation of forget instructions and thus it cannot fully explain list-method directed forgetting (Sahakyan et al., 2013). These same theorists argue instead that directed forgetting is the result of two-factors; a meta-cognitive belief in the ability to forget the items and a resulting selection of any specific forgetting strategy ranging from thought suppression to thinking of substitute thoughts (Sahakyan et al., 2013). It is thought that this decision to utilize a cognitive strategy to forget the list results in a mental context change which weakens any context clues connected to the forget list. This idea is more specifically defined as the context account of directed forgetting wherein participants preferentially expend effort to connect mental context clues to the list they are told to remember as opposed to the forget list, allowing for easier recall of said items relative to the forget list (Hanczakowski et al., 2012). Overall, while the exact mechanism is clearly disputed it is typically assumed that list-method directed forgetting is the result of some form of intentional forgetting operating during memory retrieval at test. While this is similar to other paradigms such as think/no-think, the level of ambiguity still surrounding the exact mechanism behind the effect makes comparison between the two difficult.

As opposed to list-method directed forgetting, item-method directed forgetting instead assigns each individual item to forget or remember conditions and presents them to participants one at a time alongside their instruction (e.g., Wylie et al., 2008). Due to this difference in methodology, item-method directed forgetting is largely examined using within-subjects designs. Contrary to list-method studies, the item-method paradigm consistently produces decreased memory for forget items relative to remember items in recall and recognition testing (Macleod, 1999).

Similar to list-method directed forgetting, the earlier explanations for item-method directed forgetting focused primarily on various forms of selective rehearsal of remember items and passive decay of memory for forget items (Bjork, 1972). However, more recent theorists have focused on active processes such as inhibition as the mechanism behind the effect, resulting in conflicting theories between supporters of an active mechanism and the passive account in the absence of an active mechanism. More specifically, the most prominent inhibition theories surrounding the effect suggest some form of active forgetting mechanism which becomes engaged when presented with forget instructions (Anderson & Hanslmayr, 2014). This is supported by studies which have observed various indicators of increased cognitive load during the presentation of forget items. For example, Fawcett and Taylor (2008) demonstrated that participants were slower to react to an attentional probe following the forget instruction relative to the remember instruction, indicating greater cognitive effort required in the forget condition. Fawcett and Taylor (2010) expanded on this and demonstrated that the probability of successfully inhibiting a response during the stop-trials of a stop-signal reaction task improved following the forget instructions in a directed forgetting task relative to the remember instructions, potentially indicating that the two processes engaged similar active inhibitory mechanisms. Researchers have also failed to find indicators of increased cognitive load during item-method directed forgetting. Most notably, pupil dilation (thought to be indicative of cognitive effort) was found to be larger following the presentation of remember items relative to the presentation of forget items (Lee, 2018; Scholz & Dutke, 2019), supporting theories such as selective rehearsal.

More specific neurological findings have demonstrated that participants engage a fronto-parietal inhibition network when presented with forget items which subsequently results in a downregulation of hippocampal activity, a crucial region for the successful encoding of information (Ludowig et al., 2010; Nowicka et al., 2011; Rizio & Dennis, 2013). Overall, there is still a significant amount of ambiguity and discussion surrounding the theory behind item-method directed forgetting. Regardless of the exact mechanism behind the item-method directed forgetting effect it is largely accepted that it mainly affects the encoding of memory, making it distinct from the think/no-think paradigm, which is thought to assess suppression of memory retrieval.

Retrieval-Induced Forgetting

Another paradigm commonly used in relation to intentional forgetting is retrieval-induced forgetting (RIF). During a typical RIF task, participants are first presented with lists of words belonging to one of several different categories (e.g., furniture-bed, furniture-table, fruits-lemon, fruits-orange). They are then prompted to remember some of the pairs when presented with the category and part of the associated word (e.g., furniture-ta___) while other categories (e.g., fruits) are not recalled at all. The assumption behind this practice being that when retrieving the target from memory participants are also inhibiting retrieval of items that may compete for attention. Items from the non-practiced categories are left out to serve as baseline items in later memory tests. There are typically two findings. First, words that are practiced (e.g., table) are recalled more accurately than both baseline and non-practiced words, replicating well known effects surrounding rehearsal (Craik & Watkins, 1973). Second, words not practiced but belonging to a practiced category (e.g., bed) are typically less likely to be successfully recalled during later tests of memory compared to practiced items and baseline items (Anderson et al., 1994). RIF is a fairly robust effect, occurring for many different types of stimuli including visual objects (Ciranni & Shimamura, 1999), facts (Anderson & Bell, 2001), long and short texts (Johnson & Anderson, 2004; Little et al., 2011), facial features (Ferreira et al., 2014), and many others. Similarly, RIF is also observed across many different types of memory testing besides free recall, including recognition (Spitzer, 2014), association tests (Gomez-Ariza et al., 2017), problem solving (Valle et al., 2019) and phonological generation (Levy et al., 2007).

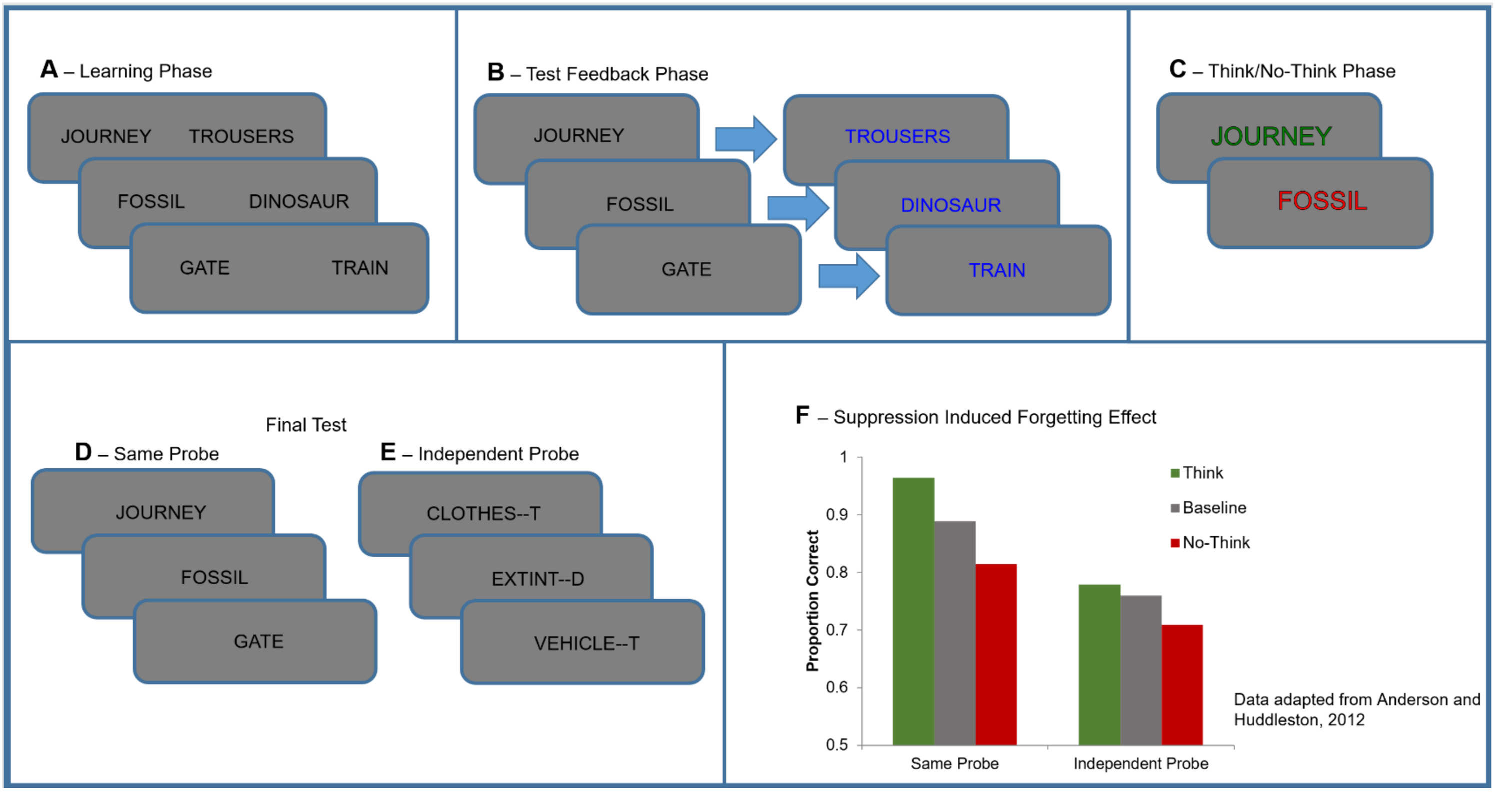
From a theoretical perspective, RIF is thought to provide evidence of inhibitory control processes which selectively target interfering items during retrieval, providing an example of intentional active forgetting (Anderson et al., 1994; Anderson & Hulbert, 2021). The typical deficits in recall for non-practiced items are also observed when using independent cues which are unrelated to the originally learned categories (e.g., using red-to____ to test for food-tomato) during recall testing (Weller at al., 2013), providing support for the idea of cognitive processes directly suppressing the words as opposed to the connection between the words and original cue. Interestingly the inhibitory processes involved in the RIF paradigm are seemingly dependent on some amount of interference during memory retrieval to begin acting, as forgetting of items is typically only observed following retrieval of other closely related stimuli. This requirement makes the RIF paradigm distinct from the think/no-think paradigm which was suggested as a method of directly studying the inhibition of recall independently of other factors such as interference.

The Think/No-Think Paradigm

The think/no-think paradigm (TNT) was originally proposed by Anderson and Green (2001) as a modified go/no-go paradigm with the goal of directly assessing retrieval suppression processes. The typical TNT procedure consists of three phases: a learning phase, a think/no-think phase, and a test phase. These phases are represented schematically in Figure 1. In the learning phase, participants are asked to learn associations between pairs of items. These materials are often word pairs (e.g., Ordeal-Roach) but can also include faces (e.g., Detre et al., 2013), images (e.g., Catarino et al., 2015), pre-recorded auditory stimuli (e.g., Cano & Knight, 2016) or even personally relevant events (e.g., Noreen & MacLeod, 2013). Once all pairs have been presented, participants are tested for those items through a series of test-feedback cycles in which they are shown each

Figure 1

Typical Procedure and Trial Events for the Think/No-Think Paradigm



Note. A) Stimuli pairs are presented next to each other during the learning phase, allowing participants to learn pair associations. B) Participants are tested for the learned stimuli pairs by providing the cue and asking for the target during the Test-Feedback phase. C) Some cue items are presented alongside a prompt to either think or not think during the Think/No-think phase. Instructions are commonly given by having the words presented either in a green font (think) or a red font (no-think) although other methods of cuing have been used depending on the chosen stimuli pairs. D) Same probe tests where the participant is given the cue item and asked for the target item E) Independent probe test where the participant is asked to recall the target when given a semantically similar word. F) Typical results of the TNT paradigm showing the SIF effect.

reminder (Ordeal) and asked to produce the associated target (Roach), with feedback provided following each response. These trials are usually repeated until participants reach either a pre-specified performance criterion or a maximum number of cycles. This is often followed by a final test wherein participants are again presented with each reminder and asked to produce the associated targets, only without feedback following their response. This test is sometimes referred to as the criterion phase and its purpose is to determine which pairs have been learned during the preceding test-feedback cycle.

In the subsequent TNT phase, participants are presented with some of the studied cues (e.g., Ordeal) again and are instructed to either (a) practice retrieving the associated target (Think trials), or (b) prevent the target from coming to mind, pushing it away if it does (No-Think trials). The remaining cues are withheld from this phase to serve as a baseline measure of memory performance during a later test. Both think and no-think items are typically presented multiple times with the presumption being that retrieval suppression becomes easier with practice and that repeated suppression of the same item should lead to a greater suppression effect (Anderson & Green, 2001).

Following the TNT phase, the test phase involves participants being asked to recall the target stimuli from the Think, No-think, and baseline conditions. Participants recall accuracy may be tested using either the same cues they had studied earlier (Ordeal-R; same-probe [SP] test) or independent cues related to the target (i.e., category) but never presented in any of the previous experimental phases (Insect-R; independent-probe [IP] test). The theory behind these different testing methods will be explored later in the paper. As one would expect (Agarwal et al., 2012; Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Roediger & Butler, 2011; Tulving, 1967), responses that were practiced during the TNT phase (Think items) are recalled better than baseline responses during the test phase. The finding of primary interest, however, is the observation that No-Think responses are recalled less often than baseline responses, a pattern which has been referred to as the SIF effect (Hertel & McDaniel, 2010).

The SIF effect has been replicated many times using word pairs (Anderson & Green, 2001; Detre et al., 2013; Hanslmayr et al., 2009) as well as autobiographical memories (Noreen & MacLeod, 2013) and visual stimuli such as pictures and face-word pairs (Catarino et al., 2015; Depue et al., 2006). However, despite multiple successful replications of the SIF effect, there remains ambiguity as to its magnitude, variability across studies, and potential moderators. Further, not all studies have produced the effect. For example, Bulevich et al. (2006) found no evidence of SIF despite replicating the original procedure by Anderson and Green (2001) with increasing detail across three experiments. A number of other studies have also failed to replicate one or more aspects of the basic behavioural effect (Detre et al, 2013; Dieler et al., 2014; Fischer et al., 2011; Hertel & Gerstle, 2003; Meier et al., 2011; Noreen & Ridout, 2016a; Noreen & Ridout, 2016b; Waldhauser et al., 2011) without a single moderator capable of explaining the varying results.

Such inconsistencies in findings have led to scepticism concerning the robustness of the SIF effect. More specifically, a review by Otgaar et al. (2019) commented that a major flaw in the TNT literature is that successful replications of the SIF effect may be limited to particular, successful research labs while failed replications are more difficult to publish. As such a review of literature published in multiple labs as well as unpublished research is critical in further establishing the credibility of SIF. Additionally, an earlier review by Raaijmakers and Jakab (2013) noted in some studies the SIF effect was present but small enough to make its practical significance questionable. As the SIF effect plays a central role in modern theories of inhibitory control, with broad applications to clinical (Catarino et al., 2015; Joorman et al., 2005; Zhang et al., 2016) and forensic psychology (Bergström et al., 2013; Hu et al., 2015), it is important that it be clearly defined both in terms of magnitude and reliability. The present thesis represents an effort to provide a comprehensive theoretical overview of the SIF effect as well as an examination of its magnitude, between-study variability, and moderators.

A previous meta-analysis conducted by Stramaccia and colleagues (2021) shared similar goals but had several notable differences which separate it from the present thesis. First and most notably, the focus of the previous meta-analysis was primarily on SIF in clinical populations, while the present thesis seeks to establish the magnitude of SIF in the general population. While Stramaccia et al. did include analyses of "healthy populations" in their paper, the samples were entirely comprised of control samples in clinical experiments who were screened for psychological disorders. As such, it is possible that these samples were in fact "healthier" than a typical population and thus may not be representative of the abilities of a typical individual. More specifically, there is evidence that executive control ability is negatively correlated with mental health in disorders such as anxiety and depression (e.g., Warren et al., 2021), and thus individuals specifically selected for scoring low on measures of mental disorders may demonstrate improved capabilities to intentionally forget relative to the average person, leading to an overestimation of the typical SIF effect. Based on this possibility, the present thesis may provide a more accurate assessment of SIF in non-clinical populations. Secondly, the previous meta-analysis only included unpublished studies in the form of theses and dissertations. Unpublished data, as highlighted in the previously discussed review by Otgar et al. (2019), are crucial for providing an accurate measure of SIF and the present thesis seeks to include a wider range of such data. Finally, due to the focus on clinical populations the previous meta-analysis included a relatively small number of studies

(25) compared to the entire TNT literature and the present thesis seeks to include a greater number of studies in order to be as comprehensive as possible. Overall, these differences highlight the remaining gaps in the literature and support the need for additional meta-analyses on the topic.

Theoretical Perspectives

In their seminal article, Anderson and Green (2001) argued that during No-Think trials participants engage an inhibitory process aimed at down-regulating the target item's neural representation to prevent it from entering awareness. This inhibition was thought to linger such that on a later test of memory the response would remain inhibited and therefore harder to recall. There are multiple areas of evidence supporting this inhibitory account. First, several studies have reported the decreased recall of target words to be independent of the cue word used to test recall, as revealed by the fact that it emerges in both SP and IP tests (Anderson & Green, 2001; Benoit & Anderson 2012; Bergström et al. 2009). Similar to in the RIF paradigm, this provides evidence that the SIF effect is a result of mechanisms inhibiting recall of the target directly as opposed to acting on the connection between components of the stimuli pair. This cue-independence occurs most consistently when participants are instructed to directly suppress the target as opposed to simply avoid thinking of the target stimuli (Benoit & Anderson, 2012). This effect of instruction on SIF provides evidence of an effort driven inhibitory mechanism, which participants can engage when instructed directly in an effort to stop thinking of specific memories. Second, individual differences in the SIF effect have been found to correlate with performance on other inhibitory tasks such as go/no-go and stop-signal inhibition (Herbert & Sutterlin, 2012; Schmitz et al., 2017), suggesting some amount of overlap between SIF and previously established inhibitory mechanisms.

Supporting these findings, research has increasingly demonstrated that retrieval suppression in the TNT paradigm invokes a fronto-parietal control network aimed at downregulating medial-temporal sites associated with general retrieval (Depue et al., 2007; Levy & Anderson, 2012) as well as sites associated with representing the unwanted memory details such as faces, spatial details and emotional content in the fusiform cortex (Gagnepain et al., 2014; Mary et al., 2020), parahipocampal place area and amygdala (Depue et al., 2007, 2010; Gagnepain et al., 2017) respectively. Specifically, successful SIF involves distinct activity patterns in the right dorsolateral prefrontal cortex (DLPFC) and ventrolateral prefrontal cortex (VLPFC; Anderson et al., 2004, 2016; Depue et al., 2007; Schmitz et al., 2017) correlated with significant downregulation of activity in the hippocampus (Depue et al., 2007; Levy & Anderson, 2012). This pattern of activity typically occurs immediately following no-think instructions, providing evidence of recruitment of cognitive resources in the pre-frontal region preceding inhibition of the hippocampus and other memory related areas. This indicates a distinct and intent driven inhibitory mechanism being recruited to suppress no-think items, which when combined with the past work supports an inhibition account of the SIF effect.

Many of these brain regions are also prominently featured in the inhibition of motor behaviours (Depue et al., 2016; Guo et al., 2018; Schmitz et al., 2017). Extending these similarities, both memory suppression and action stopping are also characterized by activity in regions within the basal ganglia such as the caudate nucleus and the putamen (Guo et al., 2018; Wiecki & Frank, 2013). Event-related potential (ERP) studies have also found connections between the SIF effect and other inhibitory mechanisms. Common inhibition related paradigms such as the go/no-go and stop-signal inhibition have been temporally linked to a significant N2 component in the frontal region of the brain (Jodo & Kayama, 1992; Kok et al., 2004), a specific pattern of neural activity associated with behavioural inhibition. The suppression of memory retrieval has also been demonstrated to elicit a larger N2 component than retrieval (Bergtrom et al., 2009). Furthermore, larger N2 amplitude during retrieval suppression is predictive of a larger SIF effect indicated by more successful suppression and subsequent forgetting (Streb et al., 2016). Memory suppression has also been shown to elicit greater power in beta waves (12 – 30 Hz) localized to the frontal brain regions, a pattern of neural oscillation shared with participants completing motor stopping tasks (Castiglione et al., 2019). These shared neural patterns suggest the SIF effect utilizes similar neurocognitive mechanisms as are employed to inhibit motor behaviours.

Not all evidence has supported the role of inhibition in producing SIF, giving rise to an alternate, interference-based account. Classical interference theories propose that conflicts between new and pre-existing information can result in the forgetting of either (Underwood, 1957). With respect to retrieval suppression, basic interference theory proposes that independently generated substitute thoughts form connections to the target stimuli similar to the cue stimuli, resulting in difficulty recalling the target at test due to a blocking effect as a result of this competition (Hertel & Calcaterra, 2005; Joorman et al., 2009; Wang et al., 2015). Further, it is argued that this blocking occurs automatically in that participants in the TNT paradigm will naturally come up with substitute thoughts to focus on when attempting to supress no-think items and that these thoughts may become associated with the cue stimuli and thus block the original target stimuli from memory via passive interference (Anderson & Green, 2001; Hertel & Calcaterra, 2005).

Further advancing the interference account, Tomlinson and colleagues (2009) proposed that an established interference model of recall, the Search of Associative Memory (SAM) model of recall (Raaijmakers & Shiffrin, 1980), can be applied to recall suppression to explain the cue-independence effect long held as evidence of inhibition as the primary account of SIF. The SAM model posits that recall consists of a sampling stage where the memory is located and a recovery stage where the memory is recalled. Applied to TNT, it is argued that individuals partially engage the recovery stage during no-think trials and the recovery of the target item can become associated with behavioural stimuli such as sitting still for a period of time. During later tests of recall, this association then causes interference and impairs recall success in both SP and IP tests, providing an alternative explanation to the inhibition account for the SIF effect (See Huber et al., 2015 for a more in-depth review).

Overall, there is evidence that inhibition and interference-related mechanisms can produce the SIF effect (Benoit & Anderson, 2012; Hertel & Calcaterra, 2005; LeMoult et al., 2010). Based on this conflicting evidence it has been proposed that thought substitution and inhibition play a role in retrieval suppression via distinct neurocognitive mechanisms (Huber et al., 2010), with multiple studies providing behavioural evidence for this dissociation in the form of the SIF effect produced by either method (Hertel & Hayes, 2015; Hulbert et al., 2016; Wang et al., 2015). Additionally, research has demonstrated that, relative to participants given direct suppression instructions ("prevent the target from coming to mind without thinking of other words"), participants instructed to use thought substitution exhibit different patterns of neural activity when engaging in retrieval suppression (Benoit & Anderson, 2012; Bergstrom et al., 2009). Overall, this provides evidence that inhibition and interference, in the form of thought substitution, are not mutually exclusive as driving mechanisms for SIF as either may play a role in retrieval suppression as defined in the TNT paradigm depending on the strategy chosen.

One perspective supporting this idea of distinct and co-occurring inhibitory and thought substitution mechanisms is the non-monotonic plasticity theory (Newman & Norman, 2010). The theory holds that neural representations of items are strengthened or weakened based on their exposure to different levels of excitability and activation. Specifically, low activation leads to no change and high levels of activation leads to a strengthening of the representation. However, neural representations experiencing moderate excitability that is comparatively less than highly excited representations instead show a weakening in connectivity. This is thought to occur due to competition between the memories, with limited resources for the maintenance and strengthening of the neural representations being selectively given to the more active memories. Applied to intentional retrieval suppression, it can be theorized that both inhibition and interference work concurrently to weaken memory strength. Namely, interference in no-think trials, caused by thought substitution, and inhibition may act to lower activity into the moderate excitability range relative to non-suppressed items where processes are then engaged to weaken the connectivity of the neural representation. This is supported by fMRI research which has demonstrated that the SIF effect follows a U-shaped curve in relation to neural activation of regions associated with neural representations of memory, with moderate activity leading to higher forgetting in the TNT task than low or high activity (Detre et al., 2013). Overall, this theory and the previously summarized studies may provide evidence to support both theoretical perspectives of TNT but there have been few direct examinations of the relative strength of the SIF effect produced using either method.

Methodological Heterogeneity in TNT

Based on these differing theoretical perspectives and various other disagreements there has been significant heterogeneity in the methodology employed in different TNT studies. Specifically, several potential moderators of the SIF effect have arisen in the recent literature and there has not yet been a direct assessment of the exact, quantified, effect of many of these variables. As such, the present thesis aims to explore several of these key moderators and the following section serves to briefly summarize some of the major factors explored in the following meta-analysis. These key moderators include recall test type, no-think instruction, target stimuli type, target stimulus emotional content, conditionalization of recall data and the previously discussed number of repetitions during the think/no-think phase of the paradigm.

Recall Test Type

As discussed above, in the final test phase of the TNT paradigm participants recall accuracy may be tested using the same cues they had studied earlier (SP test) or independent cues related to the target (e.g., a category the word belongs to) that were not presented in any of the previous experimental phases (IP test; Anderson & Green, 2001). The SP test was designed to emulate the original method through which participants learn the stimuli pairs, providing the same probe used in the initial learning phase as a prompt for recall. The IP test was proposed as a method of studying the suppression of the target item independent of its relation to the reminder cue (Anderson & Spellman, 1995; Anderson & Green, 2001; MacLeod, 2007). More specifically, the IP testing method is thought to provide a measure of intentional retrieval suppression of the target item while accounting for alternative explanations for the SIF effect, namely interference. As the interference perspective commonly holds that SIF is the result of a blocking effect disrupting the connection between the cue and target items (Hertel & Calcaterra, 2005; Joorman et al., 2009; Wang et al., 2015), changing the prompt for recall to a new item with no previous connections theoretically allows for the testing of memory for the target exclusively. While there is some concern as to whether IP memory tests are truly independent from the original stimuli pair (see Huddleston & Anderson, 2012 for a review) as well as whether interference can also explain cue-independent forgetting (Tomlinson et al., 2009), they play an important role in furthering our understanding of how recall suppression acts to weaken memory strength by isolating its effect on the target stimuli.

Overall, IP tests are often characterized by a greater variability in recall performance relative to SP tests due to the loss of context clues from using a new recall cue (Anderson & Spellman, 1995). In the context of TNT specifically, IP testing commonly results in lower overall accuracy of no-think, think and baseline items as a whole, resulting in a smaller but still present SIF effect (e.g., Anderson & Huddleston, 2012; Anderson & Green, 2001; Noreen & de Fockert, 2017). As such, it is expected that IP testing will produce a smaller and less reliable SIF effect across the literature relative to SP tests.

No-think Instruction

As mentioned in the theoretical discussion of the TNT paradigm, another important distinction that has arisen over time in the TNT literature is in the type of instruction given to participants concerning how they should prevent no-think items from coming to mind during the TNT phase. While the original paper by Anderson and Green (2001) did not provide a specific strategy for participants to use, the previously discussed developments in the theoretical underpinnings of retrieval suppression have led to two prevalent retrieval suppression strategies which are prompted by specific instructions: (1) "thought substitution" instructions, in which participants are told to suppress retrieval of the no-think response by blocking it out with another thought (e.g., "when given Ordeal think of House instead of Roach"; Hertel & Calcaterra, 2005). Participants may be given specific substitutes to use or left to generate their own; and (2) "direct suppression" instructions, in which participants are told to directly suppress the no-think response without replacing it with any other thoughts (e.g., Bergström et al., 2009). Based on the theory surrounding these distinct sets of instructions it is possible that participants may engage in unique mechanisms to prevent the retrieval of the target item depending on the instructions given. Specifically, those told to use thought substitution are promoting the creation of new connections between the cue item and non-target items that later interfere with recall (Hertel & Calcaterra, 2005; Joorman et al., 2009; Wang et al., 2015). Conversely, participants in the direct suppression condition are specifically told not to use thought substitutes and thus are more likely to engage in cognitive processes such as direct inhibition of the target item memory (Anderson & Green, 2001).

As previously highlighted, both thought substitution and direct suppression instructions are successful in producing SIF (Bergstrom et al, 2009; Hertel & Hayes, 2015; Hulbert et al., 2016; Wang et al., 2015), suggesting that both types of instruction provide a valid method of producing the effect. The relative strength of the SIF effect produced by the two methods of instruction has received little attention in the existing literature and thus remains a significant source of ambiguity in the TNT literature. For the purpose of the present thesis, it is expected that participants who receive more specific instructions, meaning either thought substitution or direct suppression, will have a better understanding of how to effectively suppress retrieval and will demonstrate a larger SIF effect than those given generic or undifferentiated instructions. A secondary area of interest is whether, when instructed to use thought substitution, participants given substitute thoughts show greater SIF than those left to generate their own. As previous research has demonstrated that participants may self-generate substitutes which are insufficiently related to the cue (Hertel & Calcaterra, 2005), it was expected that participants who are given substitutes to use would demonstrate a stronger SIF effect than those who were not.

Repetitions of Stimuli in the Think/No-Think Phase

In their seminal article, Anderson and Green (2001) implied that retrieval suppression may be a cumulative process which becomes successively stronger with each no-think trial during the think/no-think phase. Expanding on this initial assumption, several studies have included multiple levels of repetitions in their analysis ranging from 2 (e.g., Hertel & Calcaterra, 2005) up to as many as 20 repetitions of no-think items (e.g., Lambert et al., 2010). Across these studies it has consistently been demonstrated that the SIF effect does increase as the number of repetitions increases, allowing for the general conclusion that intentional forgetting is more successful given an increased opportunity to engage in it (Depue et al, 2006). Overall, these findings support the assumption that retrieval suppression is something that takes a considerable amount of effort to achieve or that the effect is additive, and that each suppression trial further diminishes the probability of that item being accessible later on. The present thesis seeks to provide a more comprehensive examination of the effect of repetition on SIF across studies, expecting that the previous results would replicate, and that the SIF effect would increase with repetition across the entire range.

Stimulus Type

The TNT paradigm has been used to study retrieval suppression across a variety of stimulus types, including variations in the cues and targets. Cue type has largely varied less than target type across studies due to the primary focus being on the item to be suppressed. However, there are studies which have used faces (e.g., Depue et al., 2006; Hanslmayr et al, 2010; Dieler, 2011) or photos of objects (e.g., Catarino et al., 2015; Nemeth et al., 2014; Sakakki et al, 2014) as the cue as opposed to words. The rationale behind this choice varies between studies but typically includes a desire to test the boundary conditions of the SIF effect, namely whether it can occur for stimuli besides words. Overall, when either faces or objects are used as the cue the TNT paradigm typically demonstrates the same SIF effect as when words are used, with similar variability in the strength of the effect observed across studies as observed with words. For example, Dieler (2011) used faces as the cue for several experiments and found that there was no SIF effect in some cases while in others there was a reversal of the typical effect with improved recall of no-think items. Contrarily, Chen et al. (2012) found a SIF effect across multiple studies while also using faces as stimuli. Studies using objects likewise at times demonstrate no SIF effect (Nemeth et al., 2014) while at others demonstrate a SIF effect (Legrand et al., 2018).

Looking at the variation in target type, the majority of studies use words as the target, but alternatives include objects (e.g., De Vito & Fenske, 2017; Chen et al., 2012), scenes (e.g., Davison et al., 2020; Depue et al., 2007) and personal memories chosen by the participants based on associations with a cue word (e.g., Noreen & Macleod, 2013; Noreen et al., 2016). In terms of the results, personal memories typically produce the SIF effect (Noreen & Macleod, 2013; Noreen et al., 2016). Using objects as the target can produce a SIF effect ranging from a small decrease in recall accuracy (De Vito & Fenske, 2017) up to a relatively larger decrease (Chen et al., 2012). Finally, studies using scenes as the target can produce a SIF effect (Depue et al., 2007) or a reversal of the typical effect with increased recall of no-think items (Dieler, 2011).

Based solely on general research examining how the encoding and recall of various types of stimuli can differ (Roediger, 1990), it is logical to question if retrieval suppression will also vary along the same parameters. However, based on the heterogeneity of the effect observed across studies in the current literature it is unlikely that the effect differs from what is observed when words are used as the cue and target stimuli. As such, it was predicted that cue and target type would have no effect on SIF, meaning that studies using non-word cues or targets would not demonstrate a different effect than those using words.

Stimulus Emotional Content

The emotional content of the chosen stimuli utilized as cues and targets can also vary between studies. Most studies examining emotional content have primarily focused on the effect of negative stimuli on clinical populations including major depressive disorder (MDD), generalized anxiety disorder (GAD) or post-traumatic stress disorder (PTSD; Joorman et al., 2009; Marzi et al., 2014; Streb et al., 2016). However, some studies have examined the effect of emotional content on SIF in more general populations. Examining the cue emotional content first, studies have examined the effect of negatively associated cues and have demonstrated both no SIF effect (Herbert & Sutterlin, 2012) and a SIF effect (Benoit et al., 2016). Studies with cues of mixed emotional content (i.e., a combination of positive, neutral, and negative) demonstrate a similar inconsistency, ranging from a reversal of the SIF effect to a SIF effect across different experiments (Lambert et al., 2010). While this pattern is based on a small number of studies it does indicate a similar overall SIF effect as observed in neutral stimuli.

Examining the emotional content of the target, there is a relatively larger number of studies that have included non-neutral stimuli. Negative stimuli are by far the most used targets of the non-neutral stimuli and they can result in a reversal of the SIF effect (Davidson et al., 2020) as well as a SIF effect (Depue et al., 2007; Legrand et al., 2018). Mixed targets represent a much smaller group of studies which produce a SIF effect (Dieler et al., 2010; Noreen et al., 2016). Finally, positive targets make up the smallest group of studies and demonstrate a SIF effect (Noreen & Macleod, 2013) or a lack of the SIF effect (Ryckman et al., 2018).

The emotional content of stimuli and the resulting recall of said stimuli are commonly linked together in theoretical perspectives, with many theorists proposing that emotional memories, particularly negative memories, are more deeply encoded and thus easier to remember. From this, it logically follows that these emotional memories would be more difficult to suppress or avoid thinking about, and thus emotional stimuli would reduce the SIF effect observed in the TNT paradigm. In support of this, research has demonstrated that emotional memories are commonly resilient to intentional forgetting using similar memory paradigms such as item-method directed forgetting (Hall et al., 2021). Further, the TNT paradigm is thought to be a laboratory analogue of everyday retrieval suppression and so it must be examined in all emotional contexts which occur in day-to-day life. However, based on the heterogeneity of effects it is difficult to differentiate the SIF effect between the two. As such, it was expected that cue and target emotional content would have no effect, meaning the SIF effect produced by using emotional stimuli as cues or targets, including positive, mixed or negative, would not vary much from that produced by neutral stimuli.

Conditionalization

The final factor examined in the present thesis which may play a role in the inter-study variability observed for the SIF effect is conditionalization of the data, meaning assessing solely the data which meets a specific condition. One of the primary issues with interpreting the results of a TNT study is the inability to differentiate between a failure to recall the target due to successful suppression of retrieval and a failure to recall the target due to a previous failure to successfully encode it. Essentially, you cannot supress the retrieval of information that was never actually learned. One way to resolve this issue in by conditionalizing the recall data on the success of initial learning, where only the recall accuracy of items successfully recalled in the learning phase are included in the final analysis (e.g., Hanslmayr et al., 2010). Data including the recall accuracy of items unsuccessfully learned in the learning phase are considered unconditionalized. The purpose of this process is to reduce the amount of noise in the data by only assessing the retrieval suppression of items that were possible to suppress. Theoretically, this process should allow for a more accurate representation of the SIF effect but it has not yet been empirically examined. As it has been established previously that there is ambiguity surrounding the strength of the true effect, this moderator may play a crucial role in furthering our understanding of SIF. As conditionalization should theoretically reduce the effect of unlearned stimuli it was expected that conditionalized data would demonstrate a larger SIF effect than unconditionalized data across studies.

The Current Research

As described in the preceding sections, there is a great deal of ambiguity and variability in our understanding of intentional forgetting in the TNT paradigm. The TNT paradigm and resulting SIF effect are an important part of the modern cognitive literature as they play a key role in our understanding of inhibitory control. More specifically, TNT has been used to establish the foundations of retrieval suppression in both theoretical cognitive research (Anderson & Hulbert., 2021) as well as more applied clinical studies (Catarino et al., 2015; Joorman et al., 2005; Zhang et al., 2016). Additionally, the SIF effect has begun to be examined increasingly in forensic research examining the possibility of criminals intentionally forgetting their memories of a crime (Bergström et al., 2013; Hu et al., 2015). To ensure the accuracy of this research and its potential real-world applications, it is critically important that we refine our understanding of not only the exact magnitude of the SIF effect but also of the many factors that may moderate the effect.

With that goal in mind, the primary aim of the present thesis was to conduct a comprehensive meta-analysis of existing empirical research examining the SIF effect as measured in general (i.e., non-clinical) samples. The present thesis seeks to provide a broader view of the overall literature by including a greater number of studies and a more in-depth analysis of key moderators than previous meta-analysis in the current literature. Overall, the present thesis seeks to examine several potential moderators of the SIF effect in an effort to establish potential causes for the notable variability of the SIF effect in the existing literature. By focusing on the key methodological factors outlined above, this meta-analysis aims to inform further theoretical debates concerning the strength and robustness of the SIF effect.

The first analyses conducted were to determine the typical SIF effect observed across all TNT studies. It was hypothesized that the typical SIF effect across all studies would range from a small to moderate effect size. Further, it was hypothesized that SP and IP testing would both produce the SIF effect, but that IP testing would result in more heterogeneity between studies and produce a relatively smaller effect size than SP testing. Following this, each moderator was examined separately to determine its influence on the SIF effect. To summarize the above, it was hypothesized that (a) stimulus type and stimulus emotional content would have little to no effect on the SIF effect; (b) conditionalization of the data would increase the magnitude of the SIF effect; (c) both direct suppression and thought substitution instructions would increase the magnitude of the SIF effect relative to no instructions; and (d) the magnitude of the SIF effect would increase as the number of repetitions increased.

Method

Search Strategy

A search was conducted of the online resources PsycINFO, PubMed, PsychNet, PsycArticles, Web of Science, ProQuest and Scopus using the Boolean search phrase: (“think/nothink” OR “think/no-think” OR “think/no think” OR “think no think” OR “TNT” OR “retrieval suppression” OR “suppression-induced forgetting” OR “suppression induced forgetting” OR “SIF”) AND (“memory” OR “recall” OR “recognition”) AND (“psychology” OR "neuroscience" OR "neuropsychology"). This search was conducted until May 2020, restricted to English-language articles, and was supplemented by (a) studies which had cited either Anderson and Green (2001) or Anderson and colleagues (2004) found using Web of Science; (b) a review of reference sections of six major review papers on TNT (Anderson & Levy, 2007; Anderson & Levy, 2009; Anderson & Huddleston, 2012; Anderson & Hanslmayr, 2014; Hu et al., 2017; Levy & Anderson, 2008); (c) communication with researchers (including multiple public social media posts as well as direct emails) who have conducted studies on TNT who may have raw or unpublished data; and, (d) studies suggested for inclusion by experts in the field.

Inclusion Criteria

Articles reporting at least one estimate of SIF as measured by SP or IP testing within a non-clinical, adult population were considered for inclusion. Articles were excluded if they 1) used only clinical samples; 2) reported no experimental data; 3) did not use the TNT task; 4) used only child and adolescent (mean age < 17) or older adult samples (mean age > 40)[1]; 5) had fewer than ten participants[2]; 6) failed to report critical aspects of the TNT paradigm (cue-target pairs, no-think phase, baseline items, within-subjects design)[3]; or, 6) did not have a final explicit memory test.

Article Screening and Selection

Duplicates (i.e., studies found in multiple databases) were identified primarily through the use of the systematic review tool Covidence supplemented by a manual review of the studies

---

[1] The criteria for older adults was based on research demonstrating that decreases in cognitive control of memory can begin at approximately 40 years of age (Salthouse, 2009).

[2] This was chosen in an effort to avoid studies containing "small study effects" wherein smaller studies report larger effect sizes and may overestimate the true effect size (Sterne et al., 2000). Remaining biases caused by small sample size were partly accounted for by standardizing the results, weighting the results based on variability. Importantly, including this criterion served to make the resulting estimates of SIF, if anything, more conservative.

[3] Authors were contacted whenever possible to determine if the missing information was collected but not included in the published paper.

identified by the initial literature search. Following this, the abstracts of all remaining studies were screened for relevance by determining if they contained a mention of SIF, TNT, memory, suppression, forgetting or other relevant topics described in the initial Boolean search phrase. Finally, a full text review was conducted on the remaining studies to determine if they met the above inclusion criteria. This process was completed by at least two team members for each study as well as consultation with my supervisor.

Data Extraction

Coding of each of the included articles was completed in consultation with my supervisor and all coding decisions were documented and discussed until a consensus emerged. Specifically, each study was coded by at least two team members and conflicts in coding were resolved via discussion. Coding focused on the recall accuracy reported in the baseline, think and no-think conditions of each experiment. In addition, the demographic information of the participant samples of each study and moderators proposed to be of theoretical importance within the TNT paradigm were also coded. These moderators included test type (IP, SP), number of repetitions in the TNT phase, no-think instructions (direct suppression, thought substitution, undifferentiated), target stimuli type (word, face, object, scene), target stimulus emotional content (positive, negative, neutral, mixed) and conditionalization of the data (conditionalized or unconditionalized).

Effect Size Calculation

Effect sizes were calculated as standardized mean change (SMC) via the *escalc* function of the *metafor* package (Viechtbauer, 2010) in *R* 3.6.3 (R Core Team, 2020). Standardization was completed using the raw score standardization method as seen in (1) and (2) (SMCR; Becker, 1988) within the *escalc* function.

$$y_i = \frac{X_B - X_{NT}}{\sigma_{NT}}$$

(1)

$$v_i = \frac{2(1 - r_i)}{n_i} + \frac{y_i^2}{(2n_i)}$$

(2)

Mean differences were calculated between mean percentage of target items recalled in experimental (no-think) and baseline conditions to assess the SIF effect. Standard deviation of the no-think condition was used for standardization in (1) due to the assumption that standard deviation of the baseline would be artificially lowered by ceiling effects which are more likely to occur in the baseline condition.[4] Both standard deviation and the correlation between conditions were required for these calculations. Whenever possible these values were taken directly from the published study or calculated from other reported values (standard error, confidence intervals, t-scores). In the cases where this was not possible and we were unable to contact the authors for clarification, the values were imputed as the average of the standard deviations or correlations available for that condition. Separate effect sizes were calculated for IP and SP tests.

Modelling Method

Following the calculation of the effect sizes, a series of Bayesian random- and mixed-effects models were generated using the *brms* 2.13.5 (Burkner, 2017, 2018) package in *R* 3.6.3 (R Core Team, 2020). Two initial models were fit to determine the average SIF effect across all studies for IP and SP tests, respectively. Following this, a series of moderator models were fit to

[4] Ceiling effects, defined in the present case as studies with accuracy in one or more conditions at or near 100%, artificially lower the standard deviation of the relevant condition, potentially inflating the estimated effect. A supplementary analysis confirmed the assumption that this would negatively impact my statistical models as using the baseline standard deviation caused several effects to increase substantially relative to the unstandardized mean difference or standardized mean difference using the no-think standard deviation, becoming outliers. Importantly, standardizing via the no-think condition served only to make our estimates, if anything, more conservative.

assess the effect of the previously described moderators on the SIF effect. These models were considered multi-level because they included random intercepts reflecting variation across studies as well as variation across samples within a given study. This approach allowed us to assume that there was some amount of variability in the effect across each study or sample which is presumably caused by heterogeneity in the methodology of the included studies. This is especially important for paradigms such as TNT which, as previously described, demonstrate significant heterogeneity in methods throughout the literature. In addition to estimating variability attributable to differences across samples and articles, this approach also estimates the average effect from a typical sample or article. Overall, this allows an assessment of not only the aggregate effect but also the amount of heterogeneity between different studies, which can then be further examined and explained using additional moderator models.

Each model was run using 4 chains with 30,000 iterations included. Each chain represented an independent version of the model, resulting in a total of 120,000 iterations. The first 15,000 iterations of each of the chains were considered as a warm-up period where the model calibrated itself and as such a total of 60,000 iterations were removed. The remaining iterations served as the posterior distribution of the model. Separate models were fit for each of the moderators described above to independently assess their effect on SIF. Models were fit and evaluated for convergence by ensuring R-hat was less than 1.01 (Gelman & Hill, 2007).

The priors utilized in these models were mildly informed through consultation with experts, including my supervisor, as reasonable assumptions made in consideration of the existing TNT literature. When comparing the magnitude of suppression-induced forgetting within a given study my prior expectations relating to the intercept of each model assumed that the average effect size in a typical sample should range between -1 and 1. I further assumed the standard deviation pertaining to random effects should range between -0.5 and 0.5 from the mean; this broadly permits the "true" effect size within any given sample to vary anywhere from -2 to 2 from the mean. Our prior for slopes within the moderator models were represented by a normal distribution centred at 0 with a standard deviation of 0.5.

Heterogeneity was quantified within each of our non-moderator models using prediction intervals (IntHoult et al., 2016), which reflect the range of probable "true" effects that would be expected should a new study be conducted like those included in the analysis. This allows us to address the true range of SIF effects reported in the current literature and reduce the previously described ambiguity surrounding the effect. I^2 was not reported as it has been shown to be interpreted inconsistently in the literature and is a less accurate representation of heterogeneity relative to alternatives such as prediction intervals (Borenstein et al., 2017). For each moderator, Bayesian *p*-values were calculated reflecting our confidence in the direction of the observed effect (e.g., a value of .95 pertaining to a positive effect would indicate that we are 95% confident that the effect is positive).

Results

Description of Studies

A flowchart detailing the search procedure and reasons for exclusions is provided in Figure 2. In total, 1495 studies were identified using the methods described above and after screening for duplicates (n = 577) and relevance to TNT (n = 667) a total of 251 studies were assessed for eligibility using full text review. After applying the inclusion criteria there was a final sample of 69 published studies and 13 unpublished studies for a total of 82 studies included, representing 106 independent effects. Study characteristics are summarized in Table 1. Included studies are noted by an * in the reference section.

Figure 2

Meta-Analysis Inclusion Flowchart

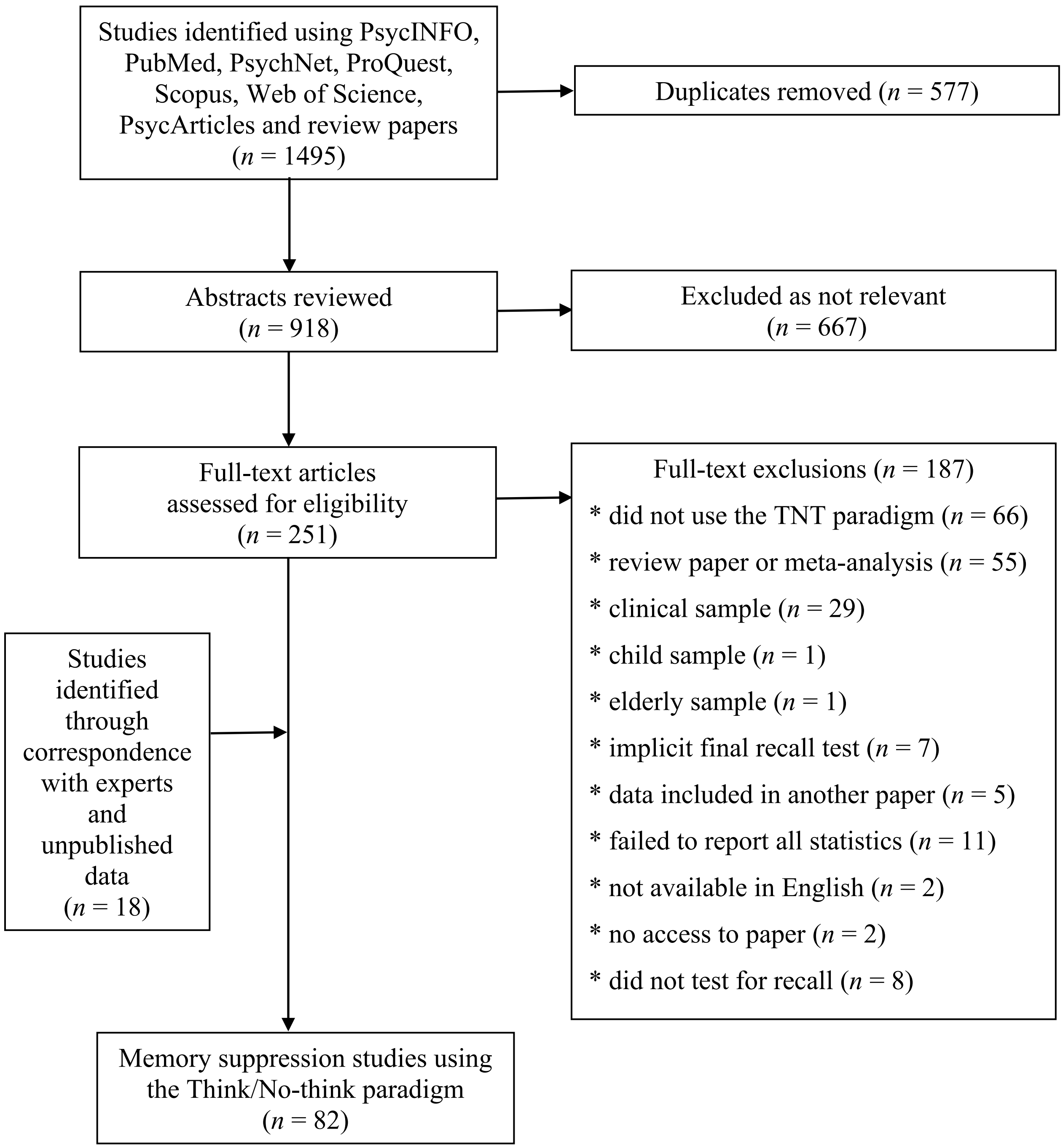


Table 1

Study Characteristics

Paper	Cue Stimulus	Cue Emotional Content	Target Stimulus	Target Emotional Content	Instruction	Repetitions	Test Type	Conditionalization
Alvez (2020)*	Word	Neutral	Word	Neutral	DS	8	SP	Both
Anderson & Green (2001; E1)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Anderson & Green (2001; E2)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Anderson & Green (2001; E3)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Anderson et al. (2004)	Word	Neutral	Word	Neutral	UD	16	SP, IP	Conditionalized
Anderson et al. (2011; E1)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Anderson et al. (2011; E2)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Attuquayefio et al. (2016)	Word	Neutral	Word	Neutral	UD	7	SP	Unconditionalized
Bell (2005; E1)	Word	Neutral	Word	Neutral	DS	1, 16	SP, IP	Unconditionalized
Bell (2005; E2)	Word	Neutral	Word	Neutral	DS	16	SP, IP	Unconditionalized
Benoit & Anderson (2012)	Word	Neutral	Word	Neutral	TS, DS	12	SP, IP	Conditionalized
Benoit et al. (2016; E1)	Word	Negative	Word	Negative	DS	12	SP	Unconditionalized
Benoit et al. (2016; E2)	Word	Negative	Word	Negative	DS	12	SP	Unconditionalized
Bergstrom (2020)*	Word	Neutral	Word	Neutral	TS, DS	16	SP, IP	Conditionalized
Bergström et al. (2007)	Word	Neutral	Word	Neutral	UD	16	SP	Conditionalized
Bergström et al. (2009)	Word	Neutral	Word	Neutral	TS, DS	16	SP, IP	Conditionalized
Bulevich et al. (2006; E1)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Bulevich et al. (2006; E2)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Bulevich et al. (2006; E3)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Cano & Knight (2016)	Word	Neutral	Word	Neutral	DS	18	SP, IP	Conditionalized
Castiglione et al. (2019)	Word	Neutral	Word	Neutral	DS	12	SP, IP	Both
Catarino et al (2015)*	Word	Neutral	Word	Neutral	DS	8	SP, IP	Both
Catarino et al. (2015)	Object	Neutral	Scene	Negative	DS	10	SP	Unconditionalized
Chen et al. (2012)	Face	Neutral	Object	Neutral, Negative	DS	10	SP	Unconditionalized
Davidson et al. (2020)	Word	Neutral	Scene	Neutral, Negative	DS	10	SP	Unconditionalized
De Vito & Fenske (2017; E1)	Word	Neutral	Word	Neutral	UD	16	SP	Unconditionalized

De Vito & Fenske (2017; E2)	Word	Neutral	Word	Neutral	UD	16	SP	Unconditionalized
De Vito & Fenske (2017; E3)	Word	Neutral	Object	Neutral	UD	10	SP	Unconditionalized
del Prete et al. (2015; E1)	Word	Neutral	Word	Neutral	TS	12	SP, IP	Conditionalized
del Prete et al. (2015; E2)	Word	Neutral	Word	Neutral	TS	12	SP, IP	Both
Depue et al. (2006; E1)	Face	Neutral	Word	Neutral, Negative	UD	5, 10	SP	Unconditionalized
Depue et al. (2006; E2)	Face	Neutral	Scene	Neutral, Negative	UD	5, 10	SP	Unconditionalized
Depue et al. (2007)	Face	Neutral	Scene	Negative	UD	12	SP	Unconditionalized
Depue et al. (2013)	Face	Neutral	Scene	Negative	UD	12	SP	Conditionalized
Depue et al. (2015)	Face	Neutral	Scene	Neutral	UD	12	SP	Unconditionalized
Dieler (2011; E1)	Word	Neutral	Word	Neutral	UD	5	SP	Unconditionalized
Dieler (2011; E2)	Face	Mixed	Scene	Neutral	UD	12	SP	Both
Dieler (2011; E3)	Face	Neutral, Negative	Scene	Neutral	UD	12	SP	Unconditionalized
Dieler et al. (2010)	Word	Mixed	Word	Mixed	UD	5	SP	Unconditionalized
Fawcett et al. (2015)	Word	Neutral	Word	Neutral	DS	12	SP	Both
Fischer et al. (2011; E1)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP	Unconditionalized
Gilchrist (2005)	Word	Neutral	Word	Neutral	UD	16	SP	Unconditionalized
Gillie et al. (2014)	Word	Neutral	Word	Neutral	UD	16	SP, IP	Unconditionalized
Guo (2015)*	Word	Neutral	Word	Neutral	UD	8	SP, IP	Both
Hanslmayr et al. (2009; E1)	Face	Neutral	Word	Neutral	DS	10	SP	Conditionalized
Hanslmayr et al. (2009; E2)	Face	Neutral	Word	Neutral	DS	5	SP	Conditionalized
Hanslmayr et al. (2010)	Face	Neutral	Word	Neutral	DS	10	SP	Both
Hellerstedt et al. (2016)	Word	Neutral	Word	Neutral	UD	8	SP	Conditionalized
Herbert & Sutterlin (2012)	Word	Negative	Word	Negative	UD	2	SP	Unconditionalized
Hertel & Calcaterra (2005)	Word	Neutral	Word	Neutral	UD, TS	2, 12	SP	Unconditionalized
Hertel & Hayes (2015; E1)	Word	Negative	Word	Negative	TS, DS	16	SP	Conditionalized
Hertel & Hayes (2015; E2)	Word	Negative	Word	Negative	DS	16	SP	Conditionalized
Hotta & Kawaguchi (2009)	Word	Neutral	Word	Neutral	UD	4, 12	SP	Conditionalized
Hulbert & Anderson (2018; E1)	Word	Neutral	Word	Mixed	UD	1, 16	SP, IP	Conditionalized
Hulbert & Anderson (2018; E2)	Word	Neutral	Word	Mixed	UD	1, 16	SP, IP	Conditionalized
Keresztes (2011)*	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Klein (2014)	Face	Neutral	Word	Neutral	DS	10	IP	Unconditionalized
Küpper et al. (2014)	Object	Neutral	Scene	Negative	DS	10	SP	Conditionalized

Lambert et al. (2010; E1)	Word	Mixed	Word	Neutral	UD	4, 16	SP	Unconditionalized
Lambert et al. (2010; E2)	Word	Mixed	Word	Neutral	TS	20	SP, IP	Unconditionalized
Lee et al. (2007; E3)	Word	Neutral	Word	Neutral	UD	8, 16	SP	Unconditionalized
Legrand et al. (2018; E1)	Object	Neutral	Scene	Neutral, Negative	DS	8	SP	Unconditionalized
Legrand et al. (2018; E2)	Object	Neutral	Scene	Neutral, Negative	DS	8	SP	Unconditionalized
LeMoult et al. (2010)	Word	Neutral	Word	Negative	TS	12	SP	Unconditionalized
Levy & Anderson (2012)	Word	Neutral	Word	Neutral	UD	12	SP, IP	Conditionalized
Liu et al. (2016; E1)	Face	Neutral	Scene	Negative	DS	4	SP	Conditionalized
Liu et al. (2016; E2)	Face	Neutral	Scene	Negative	DS	4	SP	Conditionalized
Liu et al. (2016; E3)	Face	Neutral	Scene	Neutral	DS	4	SP	Conditionalized
Mecklinger et al. (2009)	Word	Neutral	Word	Neutral	UD	16	SP, IP	Conditionalized
Meier et al. (2011; E1)	Word	Neutral	Word	Neutral	UD	10	SP	Unconditionalized
Meier et al. (2011; E2)	Word	Neutral	Word	Neutral	UD, TS	10	SP	Unconditionalized
Murray et al. (2015; E1)	Word	Neutral	Word	Neutral, Negative	UD	10	SP, IP	Conditionalized
Murray et al. (2015; E2)	Word	Neutral	Word	Neutral, Negative	DS	10	SP, IP	Conditionalized
Nemeth et al. (2014)	Object	Neutral	Word	Neutral	UD	8	SP	Unconditionalized
Nørby et al. (2010)	Word	Neutral	Word	Neutral, Negative	UD	8, 16	SP	Unconditionalized
Noreen & de Fockert (2017; E1)	Word	Neutral	Word	Neutral	DS	16	SP, IP	Unconditionalized
Noreen & de Fockert (2017; E2)	Word	Neutral	Word	Neutral	DS	16	SP, IP	Unconditionalized
Noreen & MacLeod (2013)	Word	Neutral	Memory	Positive, Negative	UD	16	SP	Conditionalized
Noreen et al. (2016)	Word	Neutral	Memory	Mixed	TS, DS	12	SP	Unconditionalized
Paz-Alonso et al. (2009)	Word	Neutral	Word	Neutral	UD	15	SP, IP	Unconditionalized
Paz-Alonso et al. (2013)	Word	Neutral	Word	Neutral	UD	15	SP	Unconditionalized
Raaijmakers (2020)*	Word	Neutral	Word	Neutral	UD	4, 12	SP	Unconditionalized
Racsmány et al. (2012; E1)	Word	Neutral	Word	Neutral	UD	8	SP	Unconditionalized
Racsmány et al. (2012; E3)	Word	Neutral	Word	Neutral	TS	8	SP	Unconditionalized
Ryckman et al. (2018)	Word	Neutral	Word	Positive, Negative	UD	16	SP	Unconditionalized
Sacchet et al. (2017)	Word	Neutral	Word	Neutral, Negative	UD	12	SP, IP	Conditionalized
Sakaki et al. (2014)	Object	Neutral	Scene	Neutral, Negative	UD	5	SP	Unconditionalized
Salamé & Danion (2007)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Schmitz et al. (2017)	Word	Neutral	Word	Neutral	DS	8	SP	Unconditionalized
Storm et al (2007)*	Word	Neutral	Word	Neutral	DS	1, 8, 16	SP, IP	Unconditionalized

Streb et al. (2016)	Word	Neutral	Word	Neutral	DS	10	SP	Both
Todorovic (2015)*	Word	Neutral	Word	Neutral	UD	8	SP	Unconditionalized
Todorovic (2016)*	Word	Neutral	Word	Neutral	UD	8	SP	Unconditionalized
Todorovic (2020)*	Word	Neutral	Word	Neutral	UD, DS	12	SP	Both
Tramoni et al. (2009)	Word	Neutral	Word	Neutral	UD	16	SP, IP	Unconditionalized
van Schie & Anderson (2017)	Word	Neutral	Word	Neutral	DS	8	SP, IP	Conditionalized
van Schie et al. (2013)	Word	Neutral, Negative	Word	Neutral, Negative	DS	12	SP	Conditionalized
Vicente (2015)*	Word	Neutral	Word	Neutral	UD	10	SP, IP	Both
Waldhauser et al. (2011)	Word	Neutral	Word	Neutral	UD	8, 16	SP	Unconditionalized
Wang et al. (2015)	Word	Neutral	Word	Neutral	TS, DS	12	SP, IP	Unconditionalized
Wessel (2020)*	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Wessel et al. (2010)	Word	Neutral	Word	Neutral	UD	1, 8, 16	SP, IP	Unconditionalized
Zhu & Wang (2020)	Word	Neutral	Word	Neutral	UD	12	SP, IP	Unconditionalized
Wang (2020)*	Word	Neutral	Word	Neutral	DS	12	SP, IP	Unconditionalized
Yuan et al. (2018)	Word	Neutral	Word	Neutral	DS	12	SP, IP	Conditionalized
Zhang et al. (2020)	Word	Neutral	Object	Neutral	DS	12	SP	Conditionalized

Note. Unpublished studies are indicated with an asterisk. Cue Stimuli = type of stimuli used as a cue; Cue Emotional content = emotional rating of the cue, mixed indicates an unspecified combination of emotional content; Target Stimuli = type of stimuli used as a target; Target Emotional content = emotional rating of the target, mixed indicates an unspecified combination of emotional content; Instruction = type of extra instruction given to participants during the think/no-think phase, TS = thought substitution, DS = direct suppression, UD = undifferentiated which indicates no additional instructions were given beyond avoiding thought of the target stimuli; Repetitions = all levels of repetitions at which the study reported accuracy; Recall Test Type = which type of tests were used to assess recall accuracy, same probe (SP) or independent probe (IP); Conditionalization = whether the reported accuracies included stimuli that were not learned in the learning phase (unconditionalized) or did not include them (conditionalized), both indicates that the study reported accuracy for the two methods separately.

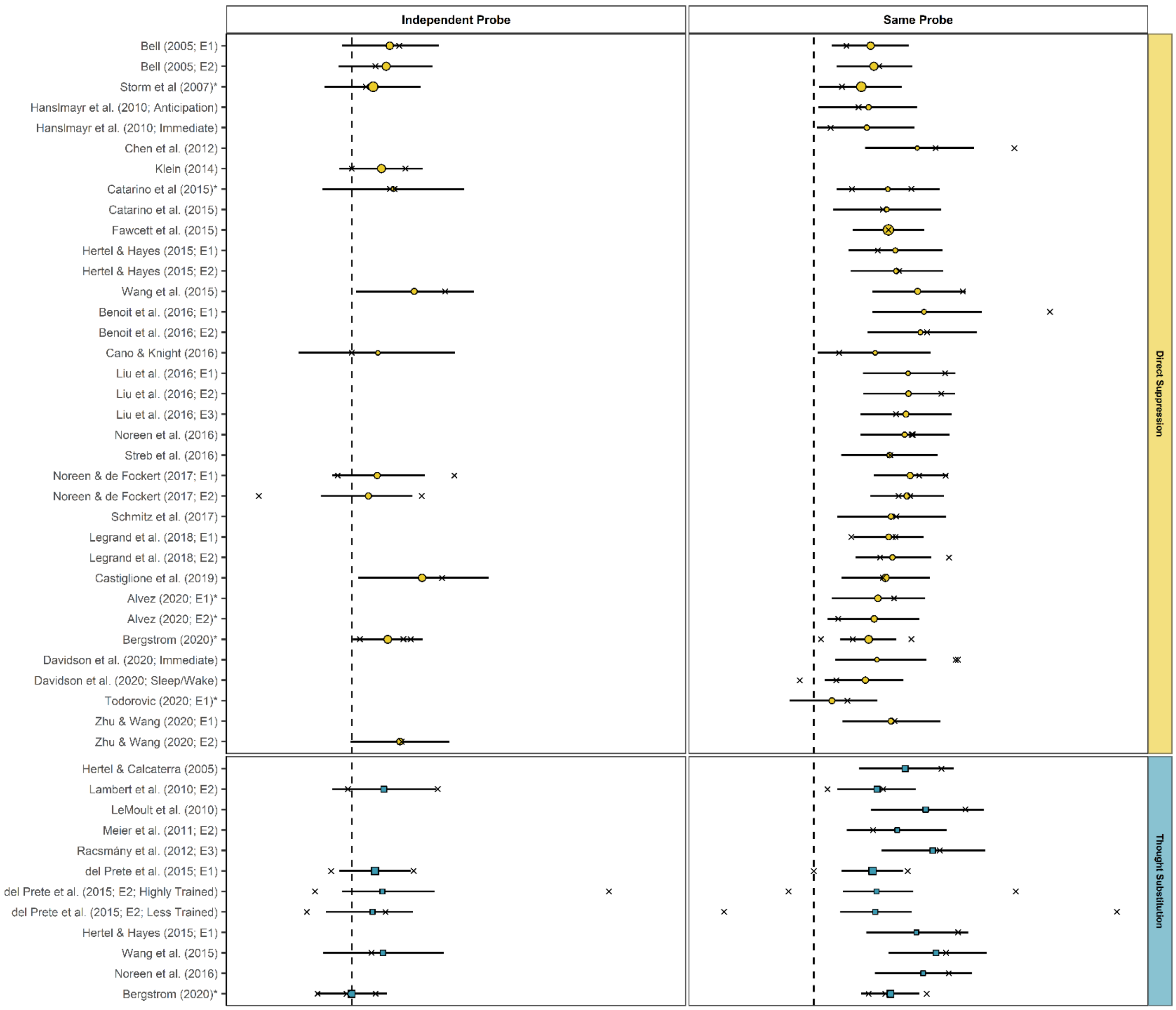
Unconditionalized Data

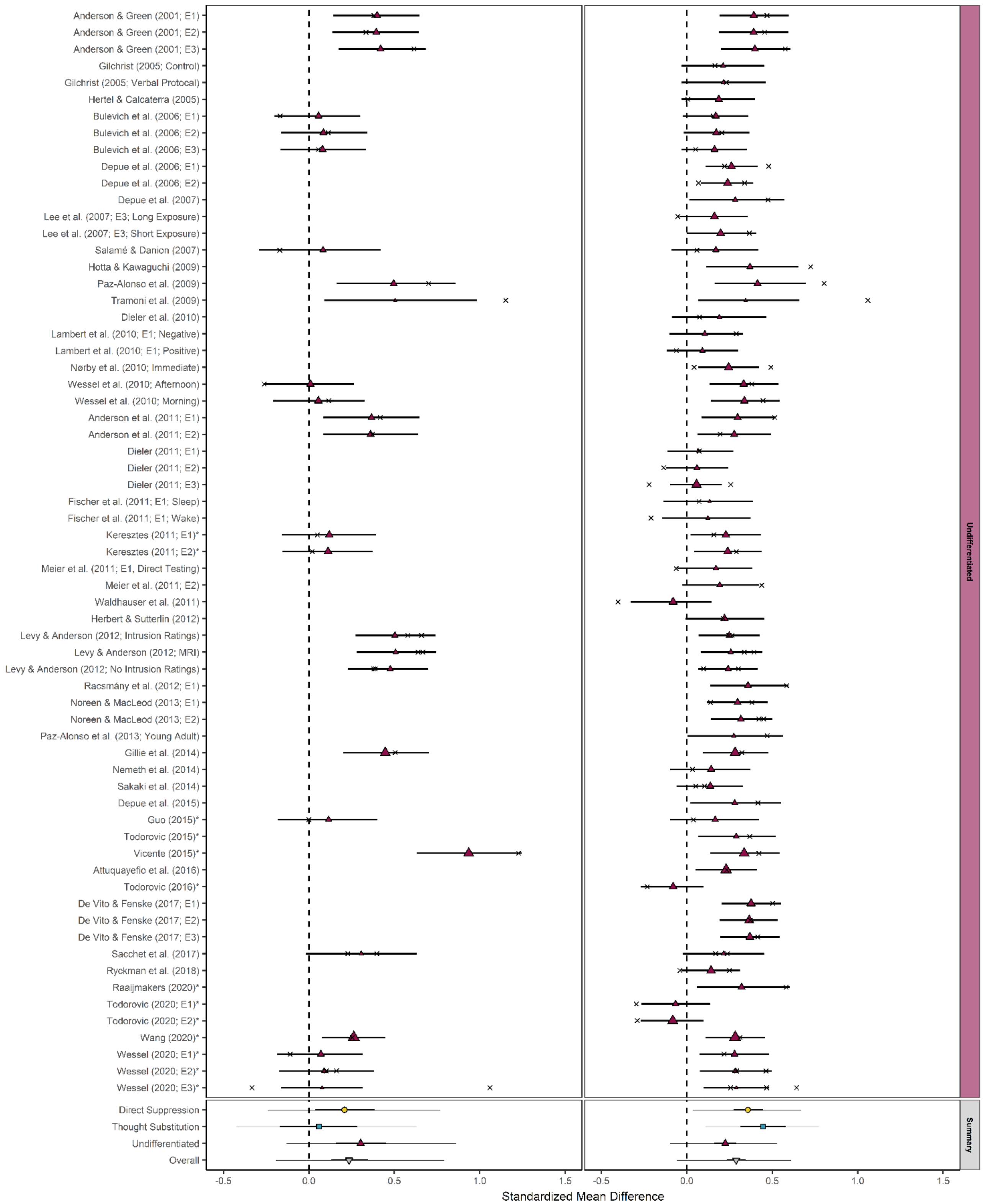
The initial analyses focused on estimating the magnitude of the SIF effect in studies reporting unconditionalized data. As shown in Figure 3, analyses were further broken down by test type (SP, IP). Specifically, due to the lack of context cues provided by changing the cue stimuli, IP tests can result in lower recall performance for both experimental and baseline items relative to SP tests. As such, it is important to analyse the two testing methods separately in order to avoid artificially obscuring the SIF effect observed across studies which utilize either method. Additionally, as summarized above it has been suggested that the two testing methods may produce results reflecting unique combinations of processes behind the SIF effect which may be useful in further informing theoretical perspectives surrounding TNT.

Analysis of the SP data revealed a small aggregate SIF effect, $SMC = 0.29$, $CI_{95\%}$ [0.24, 0.34], with broad heterogeneity in effect size across studies as evidenced by the prediction interval, $PI_{95\%}$ [-0.06, 0.61]. Analysis of the IP data demonstrated a similar aggregate SIF effect, $SMC = 0.24$, $CI_{95\%}$ [0.13, 0.35], with broader heterogeneity across studies as evidenced by the prediction intervals, $PI_{95\%}$ [-0.20, 0.79]. As the prediction interval for the SP data included small negative values it can be stated that while most studies using methods similar to those included in the present analysis may be expected to demonstrate decreased recall of no-think items, it is also possible that a small portion of studies may observe roughly no effect or even a very slight improvement in recall accuracy depending on their methods. Comparatively, the IP data demonstrated greater heterogeneity in effect size across studies, indicating that a portion of studies using similar methods may show improved recall of the no-think items at test depending on their chosen methods.

Figure 3.

Magnitude of SIF for Unconditionalized Data Produced by IP and SP Tests as a Function of Instruction Given to Participants





Note. (Yellow Circles: Direct Suppression, Blue Squares: Thought Substitution, Red Triangles, Undifferentiated). Symbols and error bars represent posterior estimates and their corresponding 95% confidence intervals. X's represent the empirical values reported in the relevant article. Symbol size is scaled to reflect relative sample size. Estimates provided in the bottom panel represent aggregate effects.

Moderators

The heterogeneity observed across studies using SP and IP testing was further analysed using moderator analyses. The overall effect of each moderator is summarized in Table 2. An insufficient number (fewer than 3) of studies were available to analyse the effects of thought substitution instruction, stimuli type and stimulus emotional content in the IP data with reliable accuracy.

No-Think Instruction. As shown in Figure 4, for studies using SP testing both direct suppression and thought substitution instructions were related to a large increase in SIF relative to undifferentiated instructions. Contrary to this finding, for studies using IP testing the effect of no-think instruction was reversed as direct suppression instructions were related to similar effect sizes as undifferentiated instructions while thought substitution instructions were related to a large decrease in SIF effect relative to undifferentiated instructions. A series of follow-up analyses were conducted to further differentiate the instruction methods in the SP data. First, an analysis comparing direct suppression and thought substitution methodologies indicated a non-credible trend of thought substitution instructions producing a greater SIF effect (difference = 0.09 [-0.03, 0.21]) with 88% confidence in the direction of the relationship. For thought substitution, it was determined that two distinct patterns of methods existed; participants were either given substitutes to use or left to generate their own. A further analysis found that studies which gave their participants thought substitutes to use demonstrated a larger SIF effect than those who did not (see Figure 5). Overall, while there was insufficient evidence to conclusively differentiate thought substitution and direct suppression, the results suggested that the inclusion of any instruction strengthens the resulting SIF effect measured by SP testing relative to undifferentiated instructions. This was further supported by a final follow-up analysis which demonstrated that studies which included either type of instruction produced greater SIF effects than undifferentiated instructions, (difference = 0.16 [0.08, 0.24]) with 100% confidence in the direction of the relationship.

Table 2.

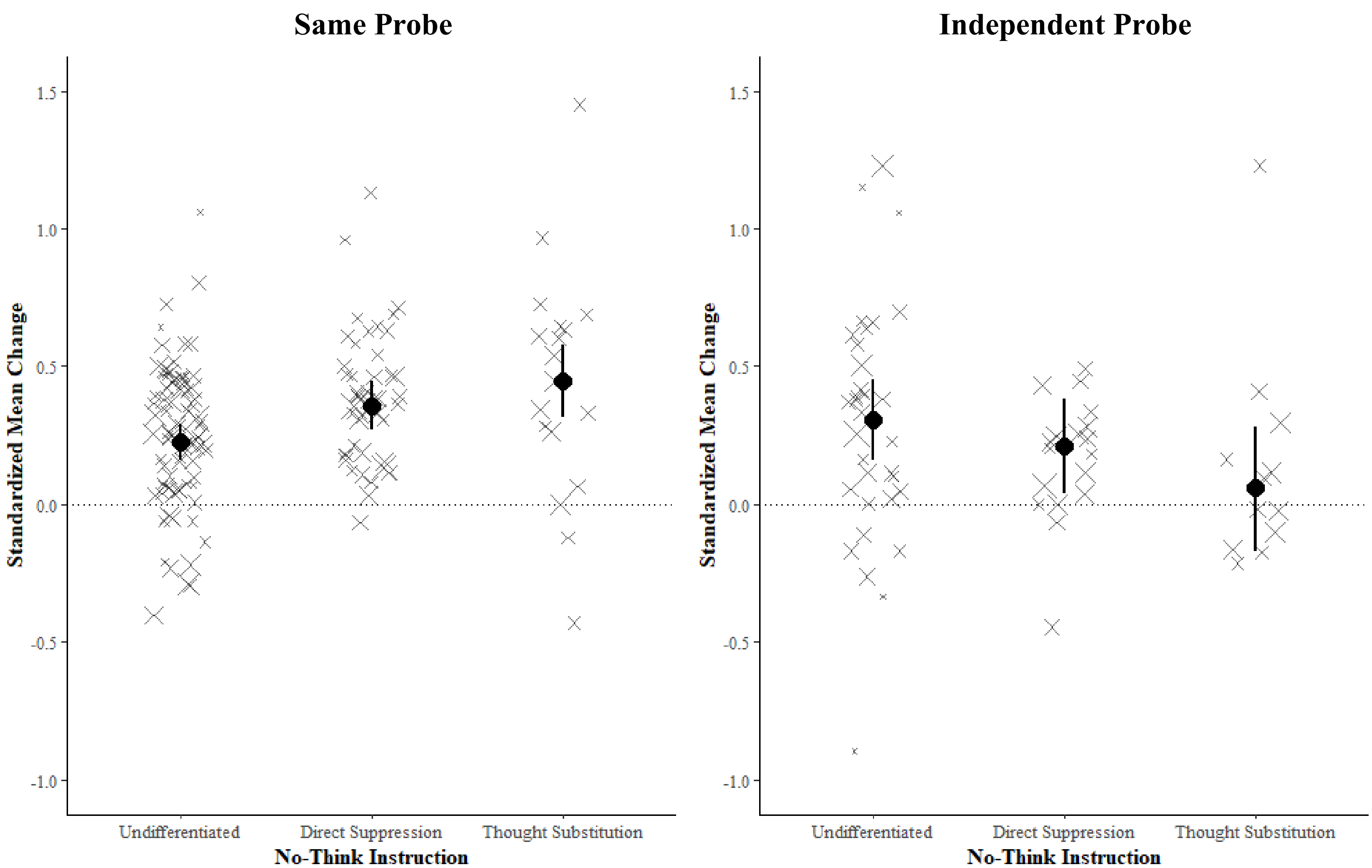
Effect of Moderators on the Magnitude of SIF for Same Probe and Independent Probe Tests

Moderator	Studies	Effects	*M*	Difference	*p*
Same Probe Test					
No-think Instruction					
Undifferentiated	42	80	0.22 [0.16, 0.29]		
Direct Suppression	23	45	0.36 [0.27, 0.45]	0.13 [0.03, 0.24]	.99
Thought Substitution	10	18	0.45 [0.31, 0.58]	0.22 [0.08, 0.36]	1.00
Substitution Instruction					
Generate Substitutes	4	12	0.31 [0.16, 0.49]		
Given Substitutes	6	6	0.58 [0.38, 0.79]	0.27 [0.00, 0.52]	.98
Repetitions	66	195	0.07 [0.04, 0.10]		1.00
Conditionalization					
Unconditionalized	59	117	0.28 [0.23, 0.33]		
Conditionalized	33	66	0.38 [0.30, 0.45]	0.10 [0.01, 0.18]	.99
Cue Type					
Words	56	118	0.29 [0.23, 0.35]		
Other	11	25	0.27 [0.14, 0.41]	-0.02 [-0.16, 0.12]	.61
Target Type					
Words	56	111	0.29 [0.23, 0.34]		
Other	13	32	0.31 [0.19, 0.42]	0.02 [-0.10, 0.15]	.64
Cue Emotional Content					
Neutral	61	130	0.30 [0.24, 0.35]		
Emotional	6	13	0.18 [0.02, 0.35]	-0.11 [-0.29, 0.06]	.90
Target Emotional Content					
Neutral	56	111	0.28 [0.22, 0.33]		
Emotional	18	32	0.34 [0.24, 0.44]	0.06 [-0.05, 0.17]	.88
Independent Probe Test					
No-think Instruction					
Undifferentiated	15	33	0.30 [0.16, 0.45]		
Direct Suppression	10	18	0.21 [0.04, 0.38]	-0.10 [-0.32, 0.13]	.81
Thought Substitution	4	12	0.06 [-0.17, 0.30]	-0.25 [-0.51, 0.02]	.97
Repetitions	27	96	0.04 [-0.01, 0.08]		.93
Conditionalization					
Unconditionalized	23	44	0.24 [0.12, 0.35]		
Conditionalized	17	39	0.33 [0.20, 0.46]	0.09 [-0.06, 0.24]	.88

Note. Studies = the number of unique articles for each moderator or each level of the moderator; Effects = the number of observations for each moderator or each level of the moderator; *M* = the aggregate estimate of SIF for each level of the moderator, in the case of continuous moderators it indicates the slope, 95% confidence interval presented in brackets; Difference = difference in magnitude of SIF between levels of the moderator, 95% confidence interval presented in brackets; *p* = Bayesian *p*-value reflecting confidence in the direction of the effect (e.g., *p* = .95 for a positive effect means 95% confidence that the effect is positive).

Figure 4.

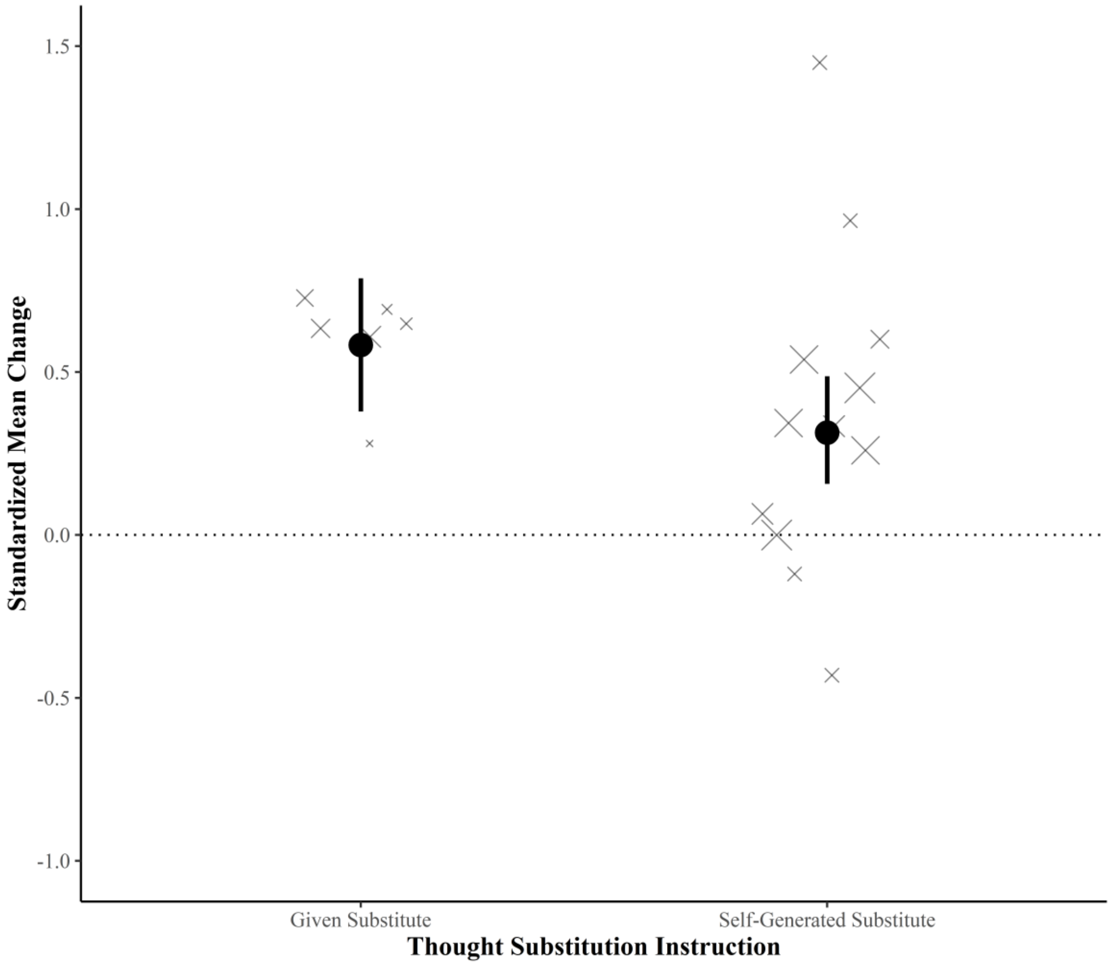
Effect of Instruction on SIF for SP and IP Testing



Note. X-axis indicates whether the instruction given to participants was undifferentiated, direct suppression or thought substitution. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect.

Figure 5.

Effect of the Type of Thought Substitution Instruction on SIF for SP Testing.



Note. X-axis indicates whether participants were given substitutes to use of told to come up with their own. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect.

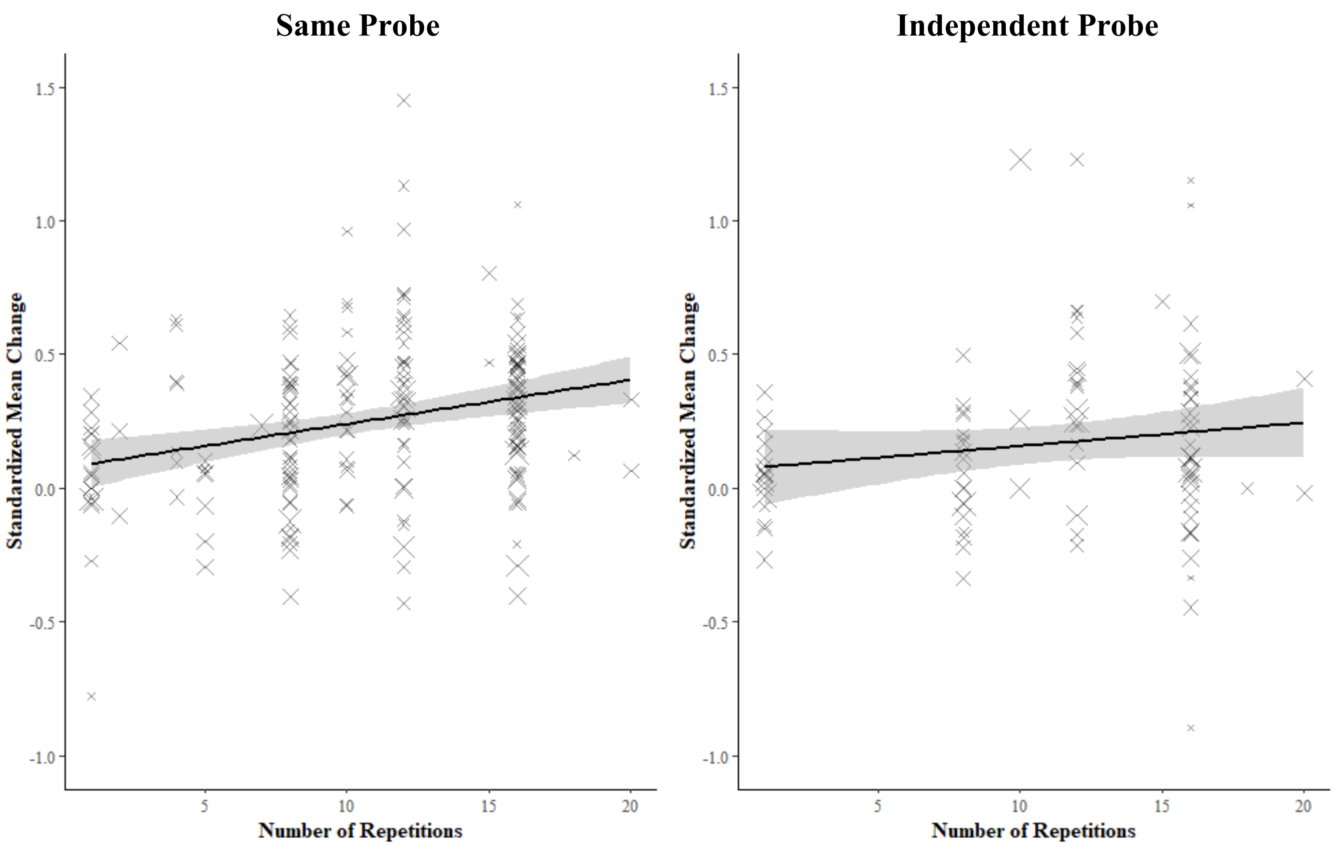
Number of Repetitions in the TNT Phase. As shown in Figure 6, there was a noticeable positive relationship between SIF and the number of repetitions, as the SIF effect increased in magnitude along with the number of repetitions. This relationship was also demonstrated in studies using IP testing although the trend was weaker, with only 93% confidence that SIF increased across repetitions, possibly due to there being far fewer effects for IP (27 studies) than SP (66 studies) and greater variability across studies.

Stimuli Type. The effects of stimuli type for SP data are summarized in Figure 7. Neither cue type nor target type were credibly associated with SIF, indicating that the TNT paradigm can produce SIF using a variety of stimuli.

Stimulus Emotional content. The effects of stimulus emotional content for SP data are summarized in Figure 7. There was a non-credible trend favouring a smaller SIF effect for emotional cues relative to non-emotional cues; however, this finding was inconsistent across studies as indicated by the broad confidence intervals of the difference, potentially due to there being only six studies using emotional cues. Further research is needed before strong conclusions can be drawn. Target emotional content had no credible relationship with SIF as the effect produced by emotional targets was similar to that produced by non-emotional targets.

Figure 6.

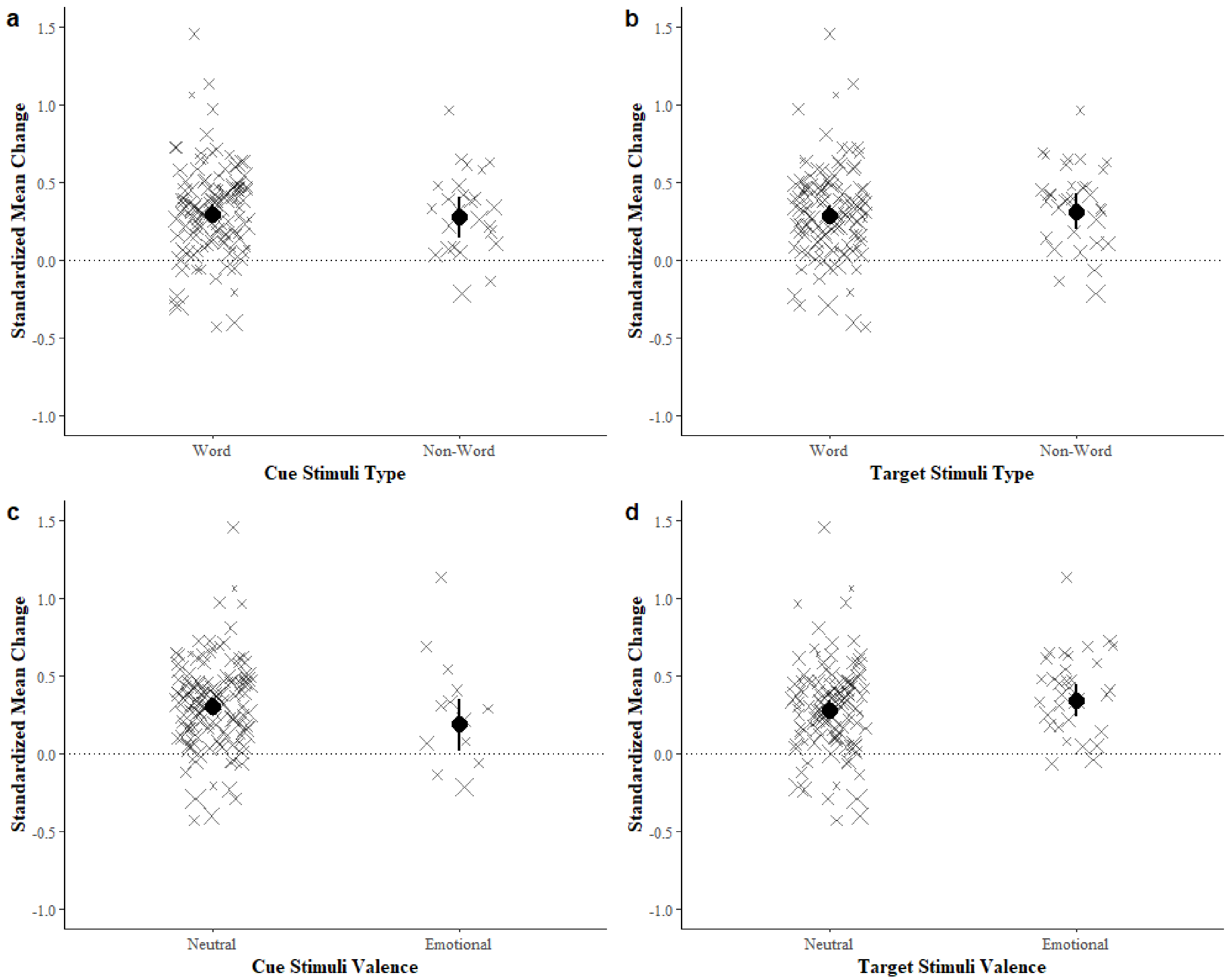
Effect of the Number of Repetitions of No-Think Items on SIF for SP and IP testing



Note. X-axis indicates the number of repetitions of items during the TNT phase. Y-axis indicates the standardized mean change. X's represent the empirical estimates of the effects of individual studies. Line indicates the regression line between repetitions and effect size. Shading around the lines indicates 95% confidence intervals.

Figure 7.

Effect of Stimuli Type and Emotional Content on SIF for SP Testing



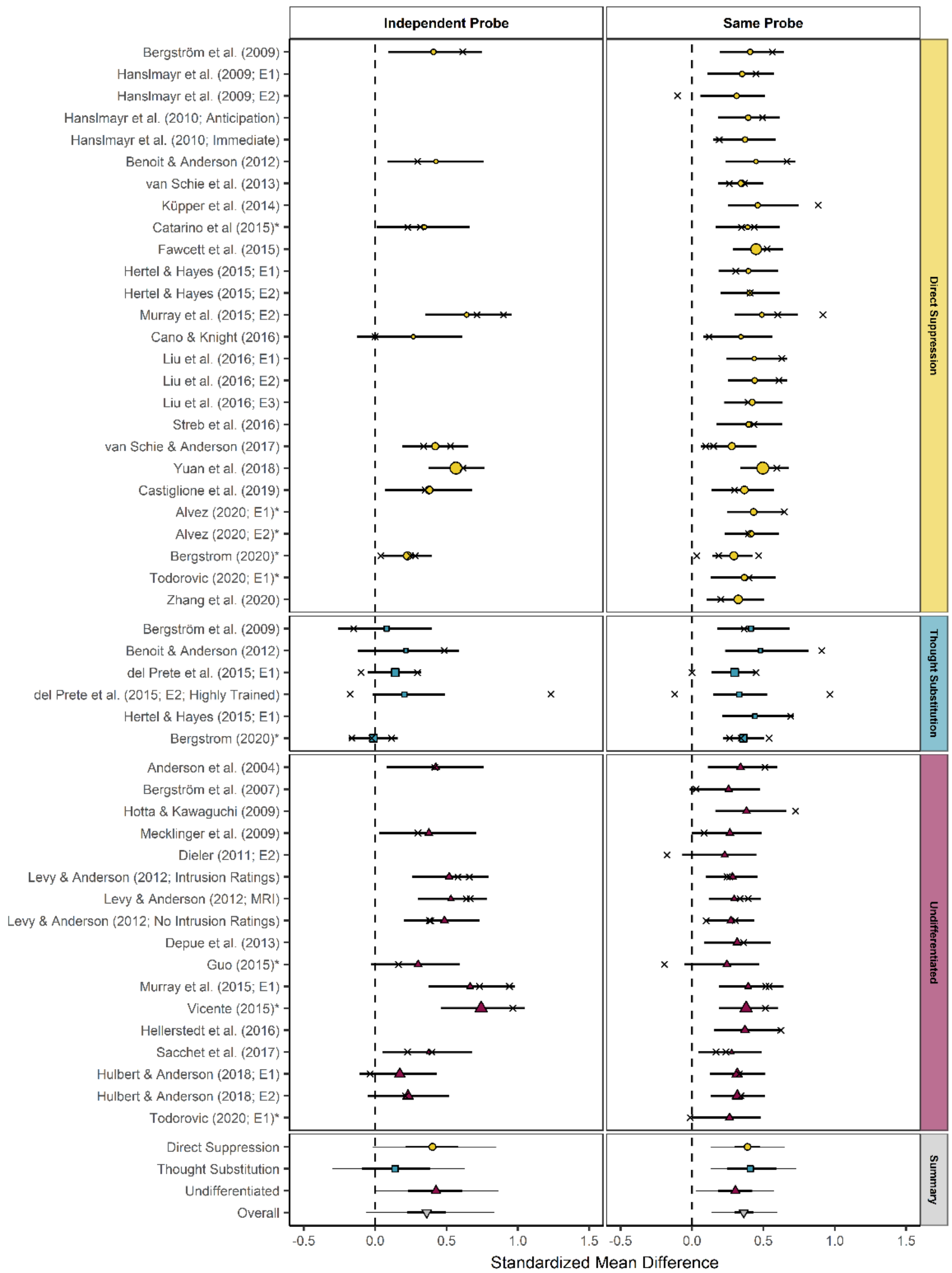
Note. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect. a) effect of cue type on SIF; b) effect of target type on SIF; c) effect of cue emotional content on SIF; d) effect of target emotional content on SIF

Conditionalized Data

A summary of the individual studies and aggregate effects for the conditionalized data is provided in Figure 8. Analysis of the conditionalized SP data revealed a small aggregate SIF effect, *SMC* = 0.36, $CI_{95\%}$ [0.30, 0.43], with broad heterogeneity in effect size across studies as evidenced by the prediction interval, $PI_{95\%}$ [0.13, 0.59]. However, the heterogeneity of the conditionalized SP data was notably less than the unconditionalized SP data as indicated by the fact that the prediction intervals did not include negative values. Analysis of the IP data demonstrated a similar aggregate SIF effect, *SMC* = 0.36, $CI_{95\%}$ [0.22, 0.50], but with broader heterogeneity across studies as evidenced by the prediction intervals, $PI_{95\%}$ [-0.06, 0.84], but still also notably less than the unconditionalized IP data. From this, it can be stated that while the average study using methods similar to those in the present analyses will demonstrate a decrease in recall accuracy for no-think items, it is also possible that a small portion of studies may demonstrate a smaller effect for SP studies or a very slight improvement in recall accuracy for IP studies.

Figure 8.

Magnitude of SIF for Conditionalized Data Produced by IP and SP Tests as a Function of Instruction Given to Participants



Note. (Yellow Circles: Direct Suppression, Blue Squares: Thought Substitution, Red Triangles, Undifferentiated). Symbols and error bars represent posterior estimates and their corresponding 95% confidence intervals. X's represent the empirical values reported in the relevant article. Symbol size is scaled to reflect relative sample size. Estimates provided in the bottom panel represent aggregate effects.

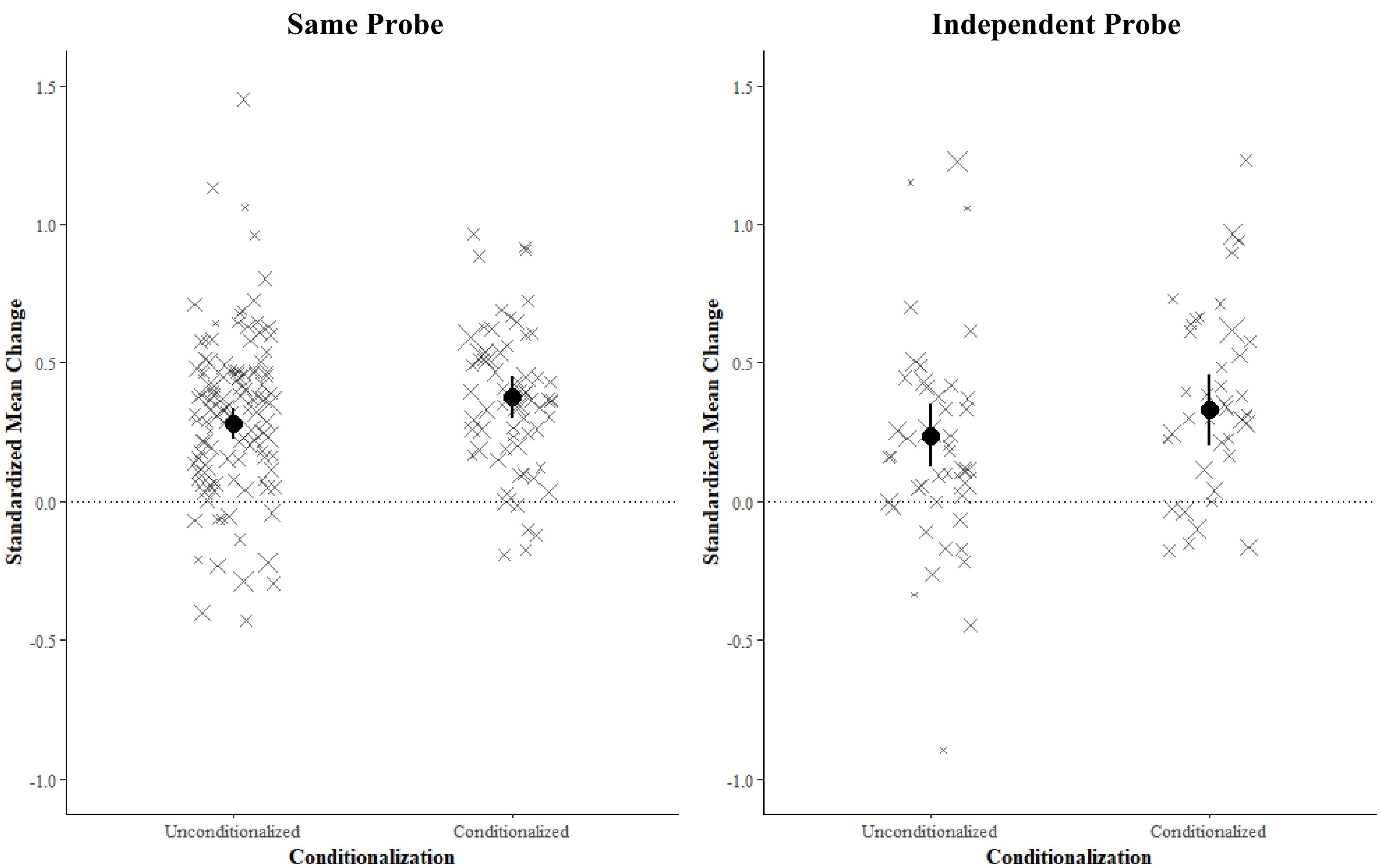
Moderators

Prior to examining the other moderators, we first compared the conditionalized and unconditionalized data to determine the overall effect of conditionalization on SIF. As summarized in Table 2, there was a trend of conditionalized data demonstrating a larger SIF effect than unconditionalized data when the two were compared across testing methods. More specifically, as shown in Figure 9 there was a trend in the SP data of conditionalization increasing the resulting SIF effect, suggesting that it accounts for some amount of noise in the data caused by unlearned items. While this trend was still noticeable in the IP data it was much weaker, suggesting that the increased variability involved in the test type may negate some of the benefits of conditionalization observed for studies using SP testing.

The effect of the remaining moderators on the conditionalized data are summarized in Table 3. Overall, these analyses largely mirrored those observed in non-conditionalized data but were less conclusive, potentially due to the lower number of studies included in the analyses. There were insufficient studies to analyse the effect of thought substitution instruction in the SP and IP data or the effect of stimuli type and cue emotional content in the IP data.

Figure 9.

Effect of Conditionalization on SIF for SP and IP testing



Note. X-axis indicates whether the data were conditionalized or unconditionalized. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect.

Table 3

Effect of Moderators on the Magnitude of SIF for Conditionalized Data

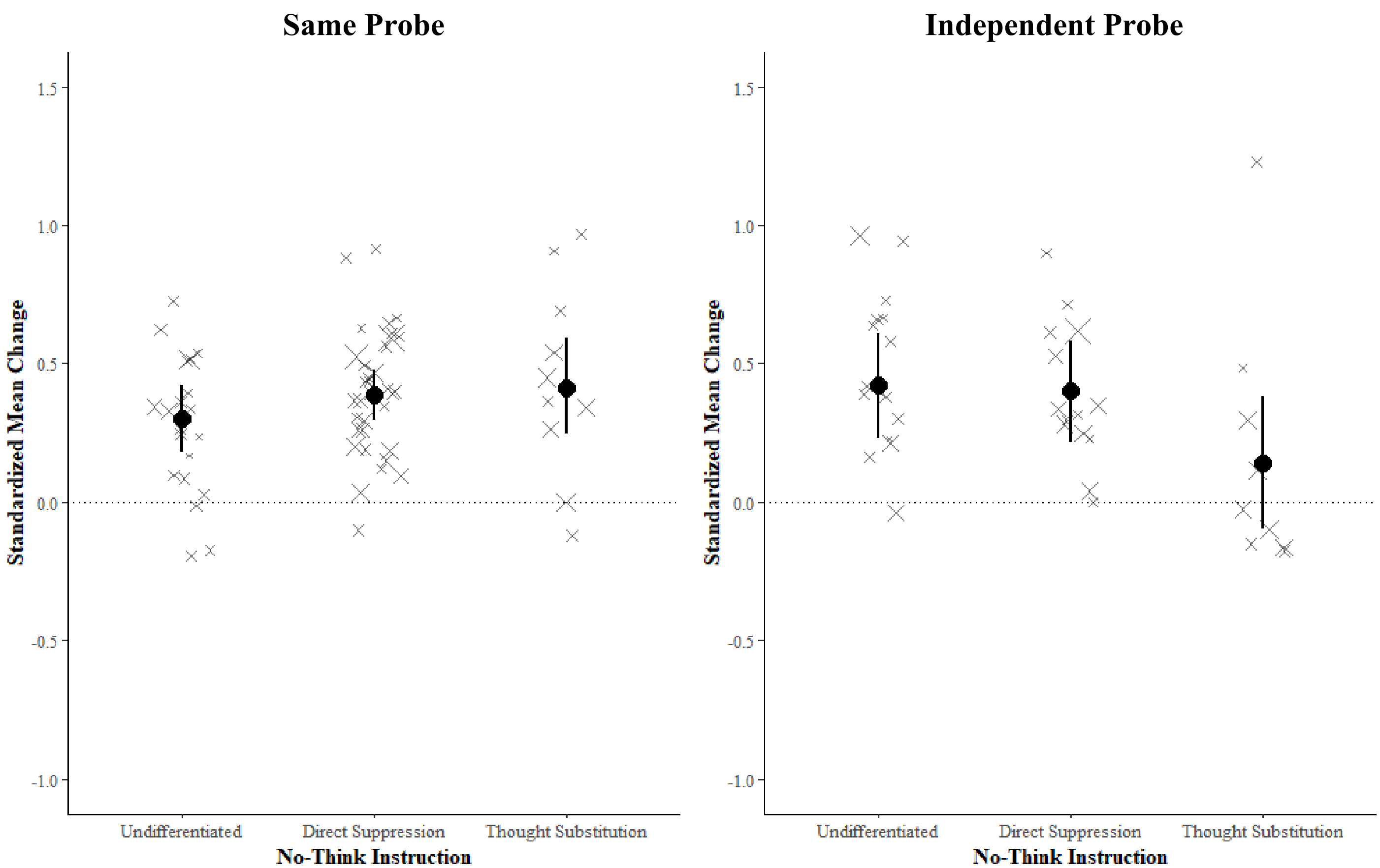
Moderator	Studies	Effects	*M*	Difference	*p*
Same Probe Test					
No-think Instruction					
Undifferentiated	14	22	0.30 [0.18, 0.42]		
Direct Suppression	20	34	0.38 [0.30, 0.48]	0.08 [-0.06, 0.24]	.87
Thought Substitution	5	10	0.41 [0.25, 0.59]	0.11 [-0.09, 0.33]	.84
Repetitions	33	69	0.00 [-0.05, 0.06]		.56
Cue Type					
Words	27	56	0.36 [0.29, 0.43]		
Other	6	10	0.35 [0.19, 0.54]	0.00 [-0.18, 0.19]	.52
Target Type					
Words	28	59	0.36 [0.29, 0.43]		
Other	5	7	0.37 [0.18, 0.56]	0.01 [-0.20, 0.21]	.54
Cue Emotional Content					
Neutral	31	59	0.37 [0.30, 0.44]		
Emotional	3	7	0.30 [0.09, 0.50]	-0.07 [-0.29, 0.14]	.75
Target Emotional Content					
Neutral	29	51	0.34 [0.27, 0.42]		
Emotional	8	15	0.43 [0.30, 0.57]	0.09 [-0.06, 0.24]	.88
Independent Probe Test					
No-think Instruction					
Undifferentiated	8	16	0.42 [0.23, 0.61]		
Direct Suppression	9	14	0.40 [0.21, 0.58]	-0.02 [-0.27, 0.23]	.58
Thought Substitution	4	9	0.14 [-0.09, 0.38]	-0.28 [-0.57, 0.01]	.97
Repetitions	17	41	-0.02 [-0.10, 0.06]		.68
Target Emotional Content					
Neutral	16	34	0.37 [0.22, 0.51]		
Emotional	3	5	0.30 [-0.01, 0.61]	-0.07 [-0.39, 0.25]	.67

Note. Studies = the number of unique articles for each moderator or each level of the moderator; Effects = the number of observations for each moderator or each level of the moderator; *M* = the mean estimate of SIF for each level of the moderator, in the case of continuous moderators it indicates the slope, 95% confidence interval presented in brackets; Difference = difference in magnitude of SIF between levels of the moderator, 95% confidence interval presented in brackets; *p* = Bayesian *p*-value reflecting confidence in the direction of the effect (e.g., *p* = .95 for a positive effect means 95% confidence that the effect is positive).

No-Think Instruction. As shown in Figure 10, for studies using SP testing both direct suppression and thought substitution instructions demonstrated a non-credible trend favouring larger SIF effects relative to undifferentiated instructions. Specifically, given that we were only 87% and 84% confident in the direction of the difference, respectively, further data is necessary to draw firm conclusions. Comparing them directly, studies using direct suppression and thought substitution instructions produce numerically similar effects (difference = 0.02 [-0.18, 0.13]) with only 59% confidence in the directionality of the relationship. Mirroring the findings observed in unconditionalized data, for conditionalized studies using IP testing the effect of no-think instruction was reversed as direct suppression and undifferentiated instructions were related to similar effect sizes while thought substitution instructions were related to a large decrease in SIF effect relative to undifferentiated instructions. While there was insufficient evidence to conclusively differentiate thought substitution or direct suppression from undifferentiated instructions in the conditionalized SP data, it was still possible that the inclusion of any instruction strengthens the resulting SIF effect measured by SP testing relative to undifferentiated instructions. A final follow-up analysis on the subject was inconclusive as it demonstrated a non-credible trend that studies which included either type of instruction produced greater SIF effects than undifferentiated instructions, (difference = 0.09 [-0.05, 0.23]) with only 89% confidence in the direction of the relationship.

Figure 10.

Effect of Instruction on SIF for Conditionalized data with SP and IP Testing



Note. X-axis indicates whether the instruction given was undifferentiated, direct suppression or thought substitution. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect.

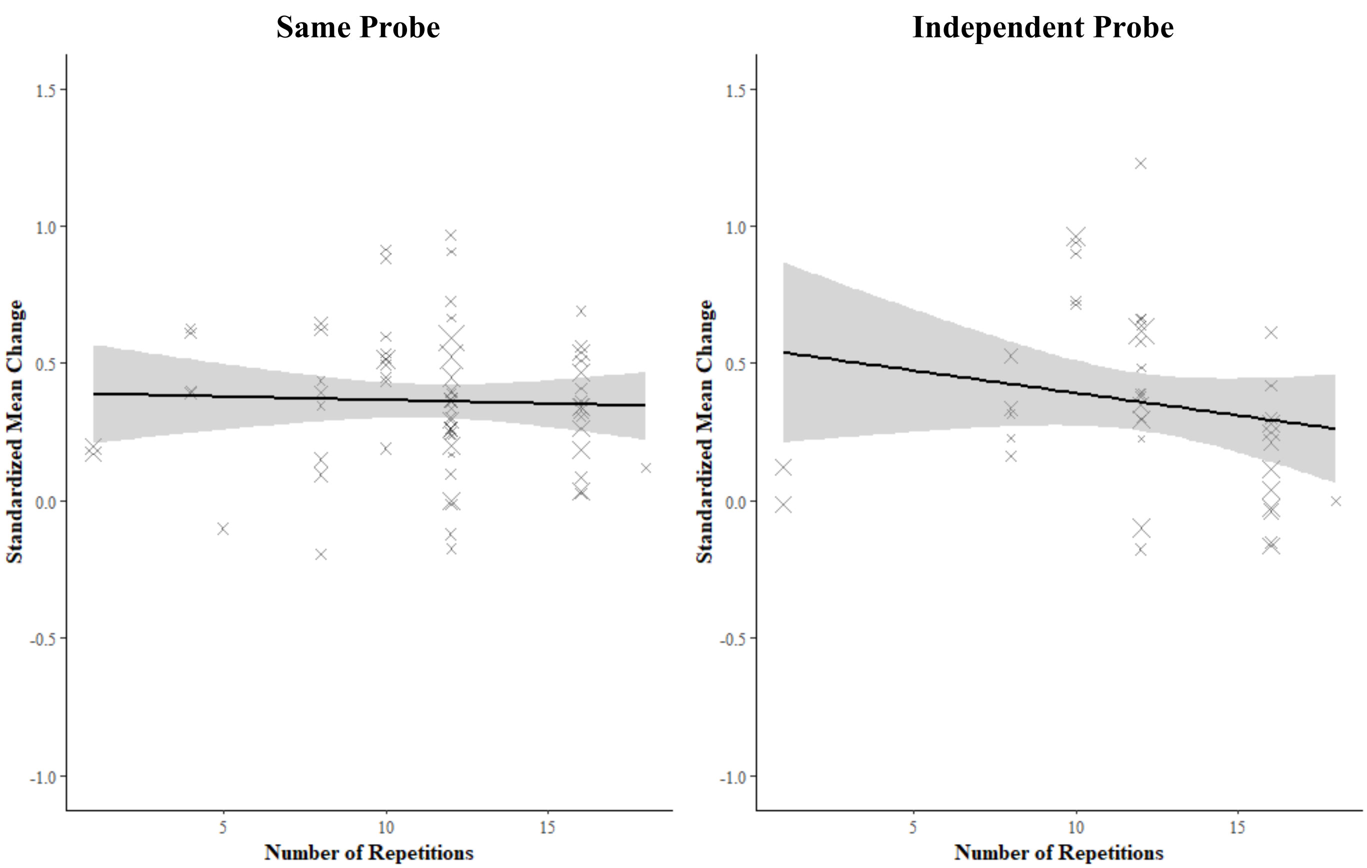
Number of Repetitions in the TNT Phase. Contrary to the unconditionalized data, in the conditionalized data there was no effect of repetition on SIF for either SP or IP testing methods. This may be partly due to the lower number of conditionalized studies, but it could also indicate that conditionalizing the data removes some of the variability that increasing repetitions may have accounted for in the unconditionalized data. As shown in Figure 11 the number of repetitions had no effect on SIF effect size in SP data and had a negative relationship with SIF effect in IP data but both trends were non-credible.

Stimuli Type. As shown in Figure 12, stimuli type did not influence the SIF effect for conditionalized SP data. Similar to in the unconditionalized analysis, this may be due to a low number of studies using non-word cues (6 studies) or targets (5 studies). From this, it can be suggested based on the existing literature that stimuli type has no effect on SIF in conditionalized data, meaning non-words produce similar SIF effects as words.

Stimulus Emotional content. As shown in Figure 12, stimulus emotional content had no influence on the SIF effect for conditionalized data. Similar to stimuli type this may be due to the low number of studies using emotional cues (3 studies) or targets (8 studies). Based on the currently existing literature it can be suggested that stimulus emotional content has no effect on SIF in conditionalized data, emotional stimuli produce similar SIF effects as neutral stimuli. However, further research into both stimulus type and emotional content is clearly needed prior to any firm conclusions being drawn. Contrary to in the unconditionalized data there was enough conditionalized data to analyse the effect of target emotional content, but as shown in Figure 13 this model also demonstrated no difference between neutral and emotional stimuli.

Figure 11.

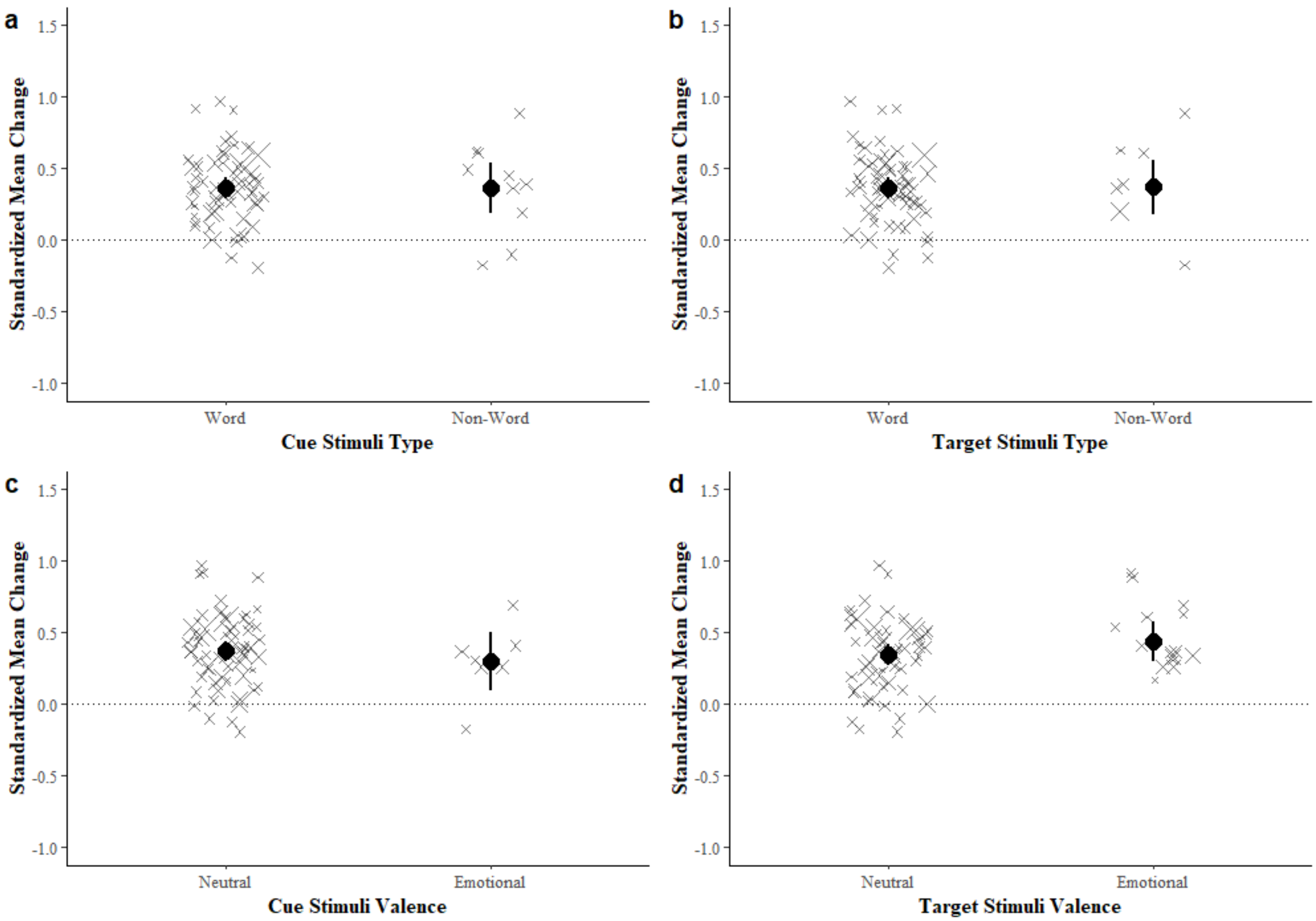
Effect of the Number of Repetitions of No-Think Items on SIF for Conditionalized Data with SP and IP testing



Note. X-axis indicates the number of repetitions. Y-axis indicates the standardized mean change. X's represent the empirical estimates of the effects of individual studies. Line indicates the regression line between repetitions and effect size. Shading around the lines indicates 95% confidence intervals.

Figure 12.

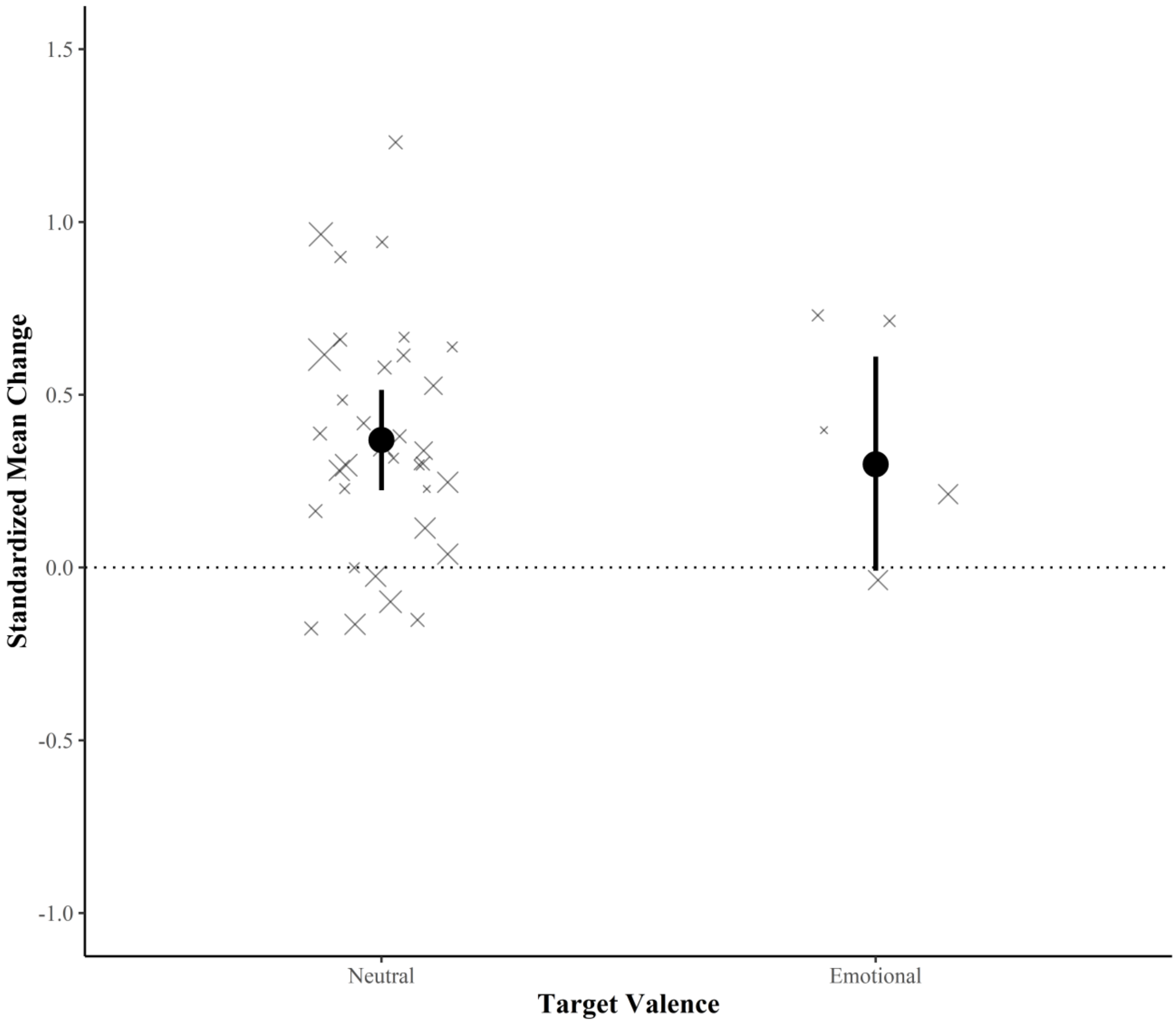
Effect of Stimuli Type and Emotional Content on SIF for Conditionalized Data with SP Testing



Note. Y-axis indicates the standardized mean change. Black circles represent the aggregate effects drawn from the posterior distribution. X's represent the empirical estimates of the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect. a) effect of cue type on SIF; b) effect of target type on SIF; c) effect of cue emotional content on SIF; d) effect of target emotional content on SIF

Figure 13.

Effect of Target Emotional Content on SIF for Conditionalized Data with IP Testing



Note. Black circles represent the aggregate effects. Y-axis indicates the standardized mean change. X's represent the effects of individual studies. Error bars indicate 95% confidence intervals. Dotted line indicates an effect size of 0 which is indicative of no SIF effect.

Publication Bias

While our analysis provides evidence of a SIF effect across all conditions, there remains the possibility that the magnitude of these effects might be driven in part by publication bias. This can occur when non-supportive findings (particularly those with relatively smaller sample sizes) are less likely to be published than supportive findings (Sutton, 2009). To evaluate this possibility, a series of regression tests were completed using the scaled standard error or sample size as a moderator, a reliable method for detecting publication bias in combination with funnel plots similar to those known as the Egger's test and Macaskill's method, respectively (Hayashino et al., 2005; Shi et al., 2017). The analyses of sample size were included in order to account for the possibility that ceiling effects may influence the results of analysis examining standard error. This was conducted separately for each combination of test type (SP, IP) and conditionalization method (conditionalized, unconditionalized).

As shown in Table 4, two of the possible eight tests demonstrated credible evidence of a relation between standard error or sample size and the aggregate effect while an additional two tests showed a non-credible trend. As shown in Figure 14 and Figure 15, standard error but not sample size credibly predicted the magnitude of SIF in the unconditionalized SP data while standard error but not sample size credibly predicted the magnitude of SIF in the unconditionalized SP data indicating a tendency for imprecise and supportive studies to be published at a greater rate than imprecise non-supportive studies. However, the failure to replicate this tendency in models examining sample size would indicate that this bias does not extend to studies with solely smaller samples. Importantly, there was less evidence of publication bias in the conditionalized data which also demonstrated a SIF effect. Neither standard error nor sample size credibly predicted the magnitude of SIF in either the conditionalized SP or IP data.

Table 4

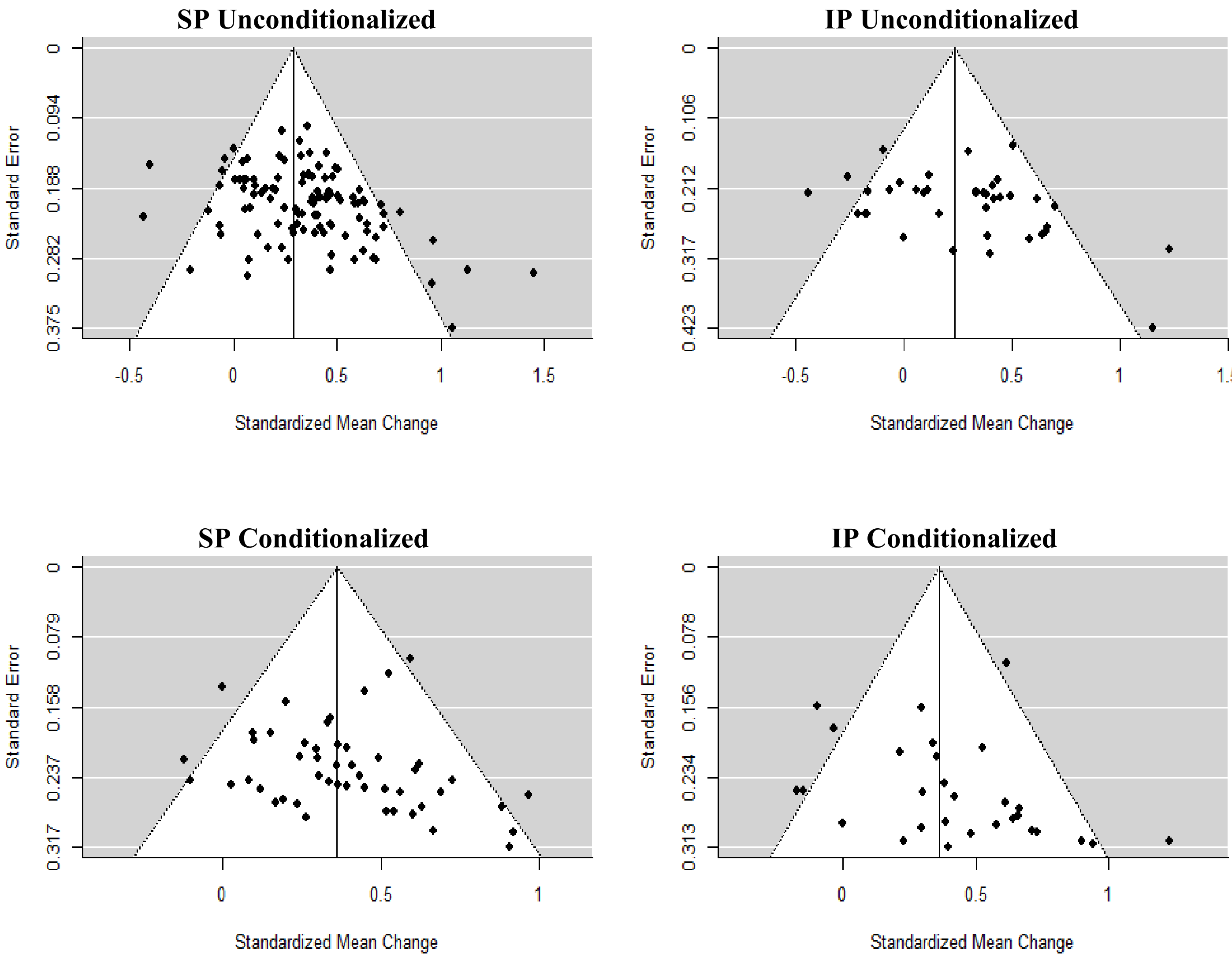
Results of the Publication Bias Analyses for all Test Types and Conditionalization Methods

	Standard Error			Sample Size		
Data	***B***	***$CI_{95\%}$***	***p***	***B***	***$CI_{95\%}$***	***p***
SP, Conditionalized	0.04	-0.02, 0.12	.89	0.02	-0.03, 0.07	.81
IP, Conditionalized	0.09	-0.03, 0.22	.94	0.01	-0.10, 0.12	.59
SP, Unconditionalized	0.11	0.06, 0.17	1.00	-0.03	-0.08, 0.01	.93
IP, Unconditionalized	0.12	0.01, 0.24	.99	-0.01	-0.11, 0.08	.61

Note. Data = the test type (same probe, independent probe) and conditionalization (conditionalized, unconditionalized) of data included in the analysis; *B* = the slope of the relationship between standard error or sample size and effect size, small sample effects are supported by positive slopes for standard error and negative slopes for sample size; *p* = Bayesian *p*-value reflecting confidence in the direction of the effect (e.g., *p* = .95 for a positive effect means 95% confidence that the effect is positive).

Figure 14.

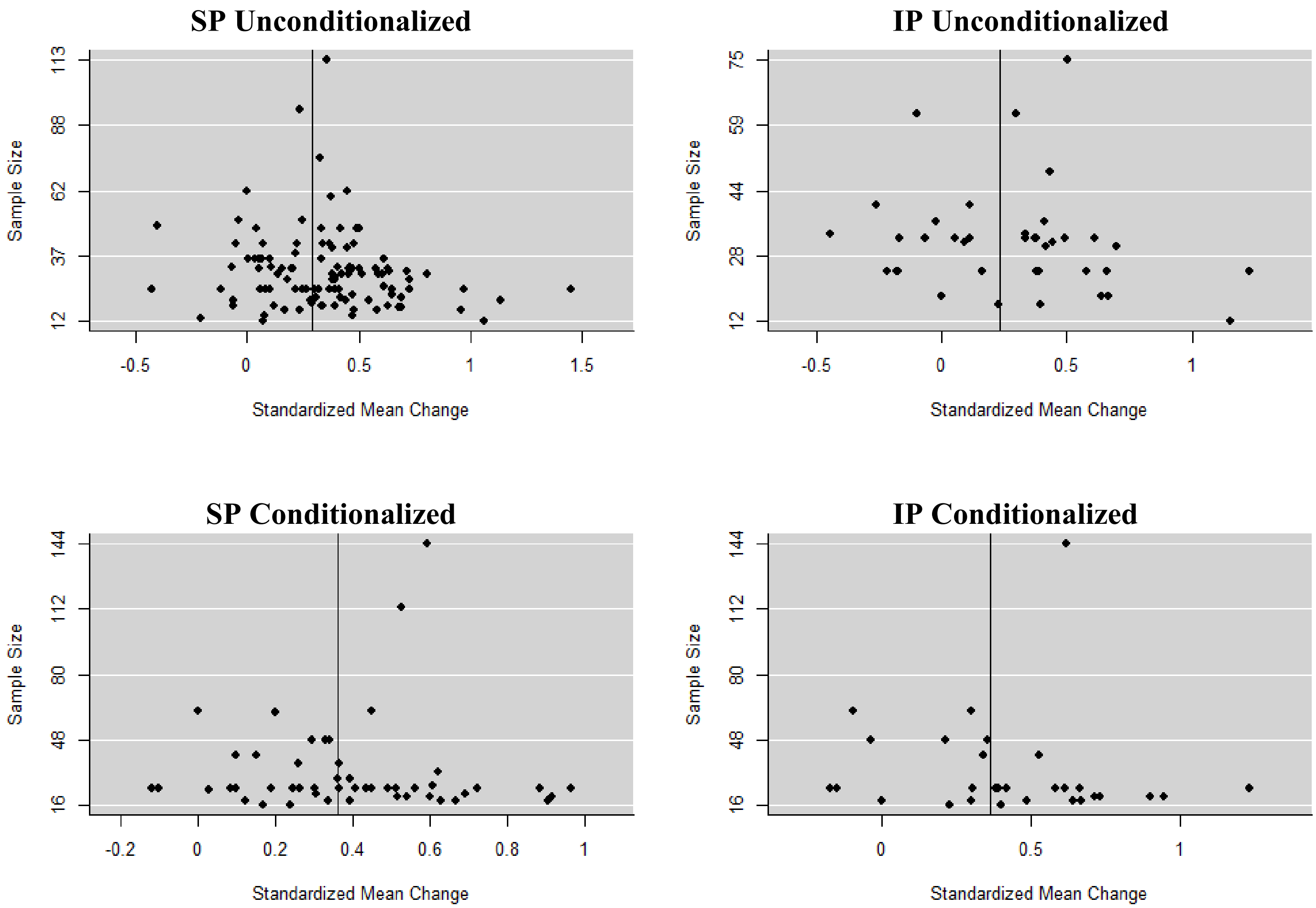
Difference in the SIF Effect Size as a Function of Test Type (SP, IP), Conditionalization (Conditionalized, Unconditionalized) and Standard Error.



Note. The y-axis of each plot reflets the range of standard error within each combination of test type and conditionalization.

Figure 15.

Difference in the SIF Effect Size as a Function of Test Type (SP, IP), Conditionalization (Conditionalized, Unconditionalized), and Sample Size.



Note. The y-axis of each plot reflets the range of sample size within each combination of test type and conditionalization.

Discussion

The current study sought to address ambiguity surrounding the SIF effect (as measured by the Think/No-Think paradigm) by (a) providing a meta-analytic synthesis of its magnitude and (b) quantifying the impact of several theoretically and methodologically motivated moderators. The former goal was driven in part by inconsistencies in the existing literature, where studies have found both a SIF effect on recall accuracy (e.g., Anderson & Green, 2001) and no SIF effect (e.g., Bulevich et al., 2006), leading some theorists to question the robustness of the effect (e.g., Otgaar et al., 2019; Raaijmakers & Jakab, 2013). Additionally, clarifying the magnitude of the SIF effect has important implications, as the TNT paradigm is increasingly employed to understand clinical disorders (e.g., Catarino et al., 2015; Zhang et al., 2016) and forensic applications (e.g., Bergström et al., 2013; Hu et al., 2015). Present models suggest SIF to be small in magnitude regardless of whether tested via SP or IP, although the effect is slightly larger for studies conditionalizing their analyses on initial learning. Prediction intervals demonstrated that it was possible, but unlikely, that future studies using similar methods would demonstrate no effect or a reversed effect using SP and IP tests, respectively, while conditionalizing the data resulted in decreased heterogeneity across studies for both SP and IP tests. Overall, the initial models confirmed the original hypothesis of TNT producing a small to moderate effect size for SIF.

The majority of the moderator models aligned with their respective initial hypotheses, with the exceptions that SIF failed to increase with the number of repetitions in the conditionalized data, that the SIF effect was smaller with thought substitution instructions in studies using IP tests, and that the effect of no-think instruction was not credible in the conditionalized SP data. More specifically, the moderator models demonstrated that the SIF effect was (a) smaller for effects measured via IP than SP tests; (b) larger for conditionalized data than for unconditionalized data (both SP and IP); (c) larger when participants were given direct suppression or thought substitution instructions instead of less specific instructions (unconditionalized SP only); (d) larger when participants were given thought substitutes to use relative to when they generated their own substitutes (unconditionalized SP only); (e) larger when the number of repetitions was higher (unconditionalized only), and; (f) unaffected by stimulus type or emotional content. These results and their implications will each be discussed in greater depth below.

Recall Test Type and the SIF Effect

Across the existing TNT literature, studies using IP tests have routinely demonstrated smaller and less consistent SIF effects than those employing the more widely used SP tests (e.g., Anderson & Green, 2001; Anderson & Huddleston, 2012; Noreen & de Fockert, 2017). This is assumed to be due to the loss of contextual information important for accurate recall (Anderson & Spellman, 1995) which results in a more varied ability to recall the target stimuli between participants. Based on this it was hypothesized that IP tests would produce a SIF effect that was relatively smaller and more heterogeneous than that produced by SP tests. This hypothesis was supported. More specifically, across both conditionalization methods the IP models showed broader heterogeneity in terms of the confidence intervals of the aggregate effect and the prediction intervals relative to the SP models. While the prediction intervals of the SP models suggested that individual studies could show no SIF effect or a very slight reversed SIF effect depending on other moderators, the IP models suggested that individual studies could show a complete reversal of the SIF effect ranging close to a small effect size. Put purely in terms of the number of studies, 24% of SP studies reported at least one negligible or slightly reversed SIF effect while 34% of IP studies reported at least one negligible or reversed SIF effect.

The theoretical implications of this result may be difficult to interpret as IP tests are a fairly controversial area of the literature with many theorists arguing not only about their validity as a testing method but also about how to best interpret their results. The IP tests were originally proposed as a method of isolating memory of the target in order to specifically study inhibition (Anderson & Spellman, 1995). This idea is partly based on the concept of cue-independence, a phenomenon observed in TNT as well as other memory paradigms wherein recall of the target item is reduced regardless of the cue used to test it, meaning recall of the target and not the cue-target relationship was impaired (Benoit & Anderson, 2012; Huddleston & Anderson, 2012). As such, SIF and other memory effects observed using IP tests are often held to be direct evidence of inhibition or direct suppression in TNT as well as in a variety of other memory paradigms (e.g., Anderson et al., 2000; Aslan et al., 2007; Macleod & Saunders, 2005; Racsmany et al., 2012). However, other researchers argue that IP tests are not truly independent and thus cannot be conclusive evidence of inhibition.

Specifically, there is evidence that covert cuing may still play a significant role in improving recall during IP tests (Camp et al., 2009), meaning that participants may still independently cue recall of the target items during testing via recall of the original cue despite there being no explicit presentation of the original cue. This process would hypothetically be vulnerable to interference effects as opposed to inhibition, implying that IP tests are not as conclusive as originally proposed. Taking another approach, Tomlinson et al., (2009) used an established framework of recall, the SAM two-stage model, to demonstrate that interference may also be a viable explanation for cue-independent forgetting, again weakening the argument that inhibition is the sole explanation for IP test results. While the relationship between IP tests and the mechanism behind SIF will be further discussed below alongside the effect of instruction, the present comparison between SP and IP testing methods conclusively shows that both are capable of producing SIF regardless of the exact mechanism behind the effect.

Conditionalization and the SIF Effect

Conditionalization is of theoretical importance because it may provide a clearer view of the magnitude of the SIF effect than the typical unconditionalized data, which is analysed in the majority of TNT studies. This idea is based on the fact that participants will naturally be unable to forget stimuli they never actually learn and including words which they incorrectly recall in the learning phase introduces random "noise" into the results, potentially weakening the observed effect. To the best of my knowledge, no prior study has directly examined conditionalization as a moderator of SIF, but the few studies reporting both unconditionalized and conditionalized results exclusively demonstrate that conditionalization is associated with an apparent increase in SIF ranging from small (del Prete, 2015; Dieler, 2011) to sizeable (Hanslmayr et al., 2011). By obtaining the raw data of several unpublished and published studies through contact with the authors and conditionalizing them, the present thesis was able to empirically examine the relationship, allowing a more thorough examination of conditionalization.

Comparing the aggregate effects of both the unconditionalized and conditionalized data there was a clear increase in the magnitude of the SIF effect when the data were conditionalized, supported by moderator models which directly compared the two groups and found that conditionalization resulted in increased SIF for studies using SP tests with a similar but weaker trend in studies using IP tests. Additionally, comparing the prediction intervals of the base models, there was a noticeable decrease in heterogeneity across studies when the data were conditionalized which supports the idea that conditionalization reduces statistical noise and increases the precision of the result. The result in the IP data lends itself to two possible explanations; (1) that IP tests and the resulting loss of contextual information (Anderson & Spellman, 1995) increases the variability of recall performance to a point in which it overpowers the positive effects of conditionalization; or (2) the analysis was underpowered due to the lower number of studies included in the IP model relative to the SP model and underestimated the true effect. There is support for either explanation as the base models examining all data for both conditionalization methods demonstrated a large increase in heterogeneity of the effect size in studies using IP testing methods relative to SP. Similarly, there being a low number of studies is in line with there still being a non-credible trend of conditionalization causing a difference in the SIF effect. Regardless, there was a conclusive effect of conditionalization on studies using SP tests which make up a large proportion of TNT research and thus a need for further interpretation.

Conditionalization of the data serves the primary function of removing potential "noise" from the results by discarding stimuli which were not learned in the initial phases of the paradigm and thus cannot be suppressed at a later point. From a statistical standpoint this should allow for greater precision in calculating the magnitude of the SIF effect, an assumption supported by the narrower prediction intervals in the conditionalized data. Based on this logic, the results of the present study would suggest that unconditionalized data typically underestimate the SIF effect, and that future studies should incorporate conditionalization into their analyses. Further, it is possible that the lack of conditionalization may contribute towards some of the failed replications of the SIF effect due to the inclusion of unlearned items (e.g., Bulevich et al., 2006). In line with this logic, the average required learning criterion for unconditionalized studies included in the present analyses was only 59%, ranging as low as 40% in some studies (Schmitz et al., 2017). Similarly, in studies where actual learning performance was reported (n = 26) participants ranged as low as 55% in terms of recall accuracy during the learning phase (Attuquayefio et al., 2016). This indicates that in some studies up to 60% of the results may have been based on unlearned stimuli pairs, introducing major amounts of noise into the data, and potentially inflating the variability of the results.

It should also be determined if there are potential confounds for the increase in SIF before making firm conclusions regarding conditionalization. If researchers who conditionalize their data also preferentially make other methodological decisions such as which no-think instruction to use SIF could be impacted positively, and such an effect may be mistakenly attributed to conditionalization. While the distribution of stimuli type, stimulus emotional content and repetitions were similar between the unconditionalized and conditionalized data, the ratio of more specific instructions (i.e., direct suppression or thought substitution) to undifferentiated instructions was higher in the conditionalized data (63%) relative to the unconditionalized data (45%). This suggests that researchers who choose to conditionalize their data also tend to include more specific instructions which may confound any improvements as a result of conditionalization. However, comparing the aggregate SIF effects for each instruction type between the unconditionalized and conditionalized data there was still an apparent increase in SIF related to conditionalization across studies using undifferentiated instructions.[5] Additionally the relationship between instruction type and SIF was weaker in the conditionalized data relative to the unconditionalized data. Overall, this seems to indicate that the effect of

[5] Follow-up analyses somewhat supported the apparent difference, showing a non-credible trend of conditionalization increasing SIF in the SP data (difference = 0.09, *p* = .89) while there was no credible effect of conditionalization in the IP data (difference = 0.04, *p* = .62).

conditionalization may be partially confounded by the effect of instruction, but that the effect of conditionalization is still notable when taking this into account. Regardless of potential confounds, conditionalization clearly plays a role in the observed SIF effect and as such, it can be concluded that TNT studies should at minimum include a comparison between the conditionalized and unconditionalized data.

No-Think Instructions and the SIF Effect

The theorized causes of the SIF effect produced by the TNT paradigm can vary wildly between different researchers, but the large majority of studies provide one of three somewhat standardized instructions to participants when explaining how to suppress thoughts of the target items in the no-think condition. Undifferentiated instructions are based on the original TNT article by Anderson and Green (2001) wherein participants were told simply to avoid thinking of the items and left primarily to decide on their own method of doing so. Based on the two most prominent theories behind SIF it is typically assumed that participants will naturally choose either direct suppression or thought substitution as their method of forgetting. However, the support for this assumption in the literature is inconsistent as some studies including post-experimental strategy questionnaires have found no pattern of chosen suppression strategy (Fischer et al., 2011) while others find a preference to use direct suppression or thought substitution (Anderson et al., 2011; Waldhauser et al., 2012). To avoid this ambiguity, some researchers may prefer to directly examine one of the two methods and thus employ either direct suppression (e.g., Hanslmayr et al., 2009) or thought substitution instructions (e.g., Hertel & Calcaterra, 2005), which theoretically prompt participants to be more selective and efficient in their strategy use.

The viability of the three distinct instruction methods is widely debated in the literature but recent research has established two key findings; (1) That direct suppression and thought substitution are not mutually exclusive as mechanisms behind SIF and can both be used effectively within the TNT paradigm to produce the effect (Bergstrom et al, 2009; Hertel & Hayes, 2015; Hulbert et al., 2016; Wang et al., 2015); and, (2) that both instruction methods produce distinct patterns of neural activity which supports that they are at least partially successful in differentiating the two underlying mechanisms (Benoit & Anderson, 2012; Bergstrom et al., 2009). With these facts in mind, the present thesis argues that it is more informative for both theoretical perspectives to establish the relative magnitude of SIF resulting from either mechanism rather than debating whether either mechanism is the cause of SIF. As such, it was assumed that all three instructions were viable methods of producing SIF and they were compared for relative strength. Examining the effects of these three different instruction methods across the literature, there was a weakening in the effect of no-think instruction between the unconditionalized and conditionalized data, specifically in studies using SP tests, but overall the trends shown in both were similar. There was however a major difference in the effect of instruction in studies using the SP testing method relative to those using the IP testing method.

Same Probe Testing

Starting with the SP data, there was a clear increase in the size of the SIF effect when participants were given the direct suppression or thought substitution instructions relative to undifferentiated instructions in the unconditionalized data. The conditionalized data contained far fewer studies and thus was less conclusive but did still show a similar trend of instruction improving SIF. Comparing the aggregate effects of each instruction across the conditionalization methods, the weakening of the influence demonstrated by instruction may have been caused by undifferentiated instructions producing a greater SIF effect in conditionalized data. More specifically, the aggregate effects of direct suppression and thought substitution studies did not increase to the same degree as the undifferentiated studies when comparing the conditionalized data to the unconditionalized. This would suggest that the more specific instructions can prompt participants to forget items with weaker encoding more effectively and that removing the influence of these items by conditionalizing based on learning performance may remove this benefit from the resulting SIF effect. It is important to note that this interaction is speculative and without further research it can only be concluded that no-think instruction had less influence in the conditionalized data.

There are several potential explanations for the effect observed in the unconditionalized data, the simplest being that giving the participants extra instructions may increase their motivation and engagement in the rest of the experiment relative to the undifferentiated instructions, resulting in larger effect sizes. Previous research has demonstrated that the level of instruction provided improves the invested mental effort and engagement of participants both in practical situations such as classrooms (Heijltjes et al., 2015) and in experimental settings (Seli et al., 2019). As SIF and many other subsets of active forgetting are commonly thought to be processes requiring effort and engagement to successfully induce forgetting (Anderson & Huddleston, 2020), it logically follows that any factor that increases engagement and effort should improve intentional forgetting. While this may partially explain the difference between instruction methods it is unclear that these more in-depth instructions such as direct suppression and thought substitution are perceived as increased engagement by participants. As such, other explanations must also be considered until direct evidence is available to support this theory.

An alternative explanation for the observed result is that by directly instructing participants to use one of the two methods as well as potentially explaining how to use them most effectively, researchers are allowing participants to focus solely on a single evidence-driven method and thus maximize its effectiveness. More specifically, cognitive resources are limited, and research has demonstrated that direct suppression and thought substitution can increase neural activity in separate regions of the brain (Bergstrom et al., 2009), implying that they engage distinct processes. As previously mentioned, compared to more specific instructions it is relatively unknown what methods participants might attempt to use when given undifferentiated instructions. Thus, participants may be engaging in a number of inefficient practices such as employing ineffective strategies (i.e., not direct suppression or thought substitution), switching between forgetting methods between trials, trying to employ multiple methods at the same time, or others which strain the capabilities of cognition and reduce the effectiveness of SIF. The other instruction methods in comparison would reduce this issue by guiding participants towards a specific method that is known to be effective.

Examining the neurological support for this interpretation, participants given the undifferentiated instruction typically demonstrate an increase of activity in the frontal region followed by a subsequent reduction of activity in the frontal and hippocampal regions (Anderson et al., 2004). Comparatively, EEG research has demonstrated that direct suppression instructions reduce activity of frontal regions followed by a reduction of activity in central and parietal regions, while thought substitution instructions produce a smaller reduction of activity in the frontal region followed by normal activation in central and parietal regions (Bergstrom et al., 2009). These results suggest that undifferentiated instructions produce activity closer to that of direct suppression but there is some overlap with the initial activity produced by thought substitution. Combined with the previously mentioned inconsistency in reported strategy use (Anderson et al., 2011; Fischer et al., 2011; Waldhauser et al., 2012), participants given the undifferentiated instructions may employ a mix of mechanisms with potentially some preference for direct suppression. This provides support for the idea that more specific instructions are related to more definitive and effective active forgetting strategy use.

It seems likely a combination of the aforementioned interpretations or similar factors improve the effectiveness of SIF when participants are given more specific instructions. Regardless, based on the results of the present study the initial hypothesis concerning no-think instruction was confirmed and both direct suppression and thought substitution instruction produced a larger SIF effect than undifferentiated instructions. While there was no hypothesis concerning a comparison between direct suppression and thought substitution instructions, an exploratory analysis found that they could not be differentiated and thus it can only be concluded that they produce a similar magnitude of SIF based on the currently existing literature.

Examining thought substitution instructions specifically, it was identified that there were two distinct methodologies employed for how substitute thoughts were generated in the literature. For example, Lambert et al. (2010) conducted a series of experiments examining the effect of emotion on SIF, telling participants in the second experiment that using alternative thoughts was an effective method for avoiding thoughts of the no-think items. Conversely, Lemoult et al. (2010) instead had participants study substitute words which they were instructed to think about instead of the original target words, providing substitute thoughts directly to the participants. These two distinct methods were examined in an exploratory analysis and it was demonstrated that participants who were given substitutes to use produced a relatively larger SIF effect than those who self-generated substitutes. This is in line with the thought substitution literature which has suggested that participants will naturally generate poor substitute thoughts which are relatively insufficient to effectively interfere with the target stimuli compared to substitutes given to them (Hertel & Calcaterra, 2005). While this result could potentially be used to argue that the difference between thought substitution and undifferentiated instructions observed in the initial moderator analysis was solely caused by studies where participants were given substitute thoughts, a situation which would never naturally occur in day-to-day life, it should be noted that the aggregate effect of studies where participants generated the substitutes (*SMC* = 0.31) was still higher than that of the undifferentiated studies (*SMC* = 0.22).

Independent Probe Testing

In comparison to the SP data, studies using IP tests demonstrated a reversal of the hypothesized effect of no-think instruction on SIF in both the unconditionalized and conditionalized data. More specifically, for IP studies undifferentiated and direct suppression instruction methods produced similar SIF effects while thought substitution instructions practically eliminated the effect. While this conflicts with the initial hypothesis that both recall test types would follow the same trend for the effect of instruction it is in line with the original purpose of the IP testing method. IP tests were initially suggested as a method of reducing the effect of thought substitution and interference to specifically examine the effect of direct suppression and inhibition (Anderson & Spellman, 1995). This is accomplished by the act of switching the cue stimuli at test to remove the potential of substitute thoughts interfering via blocking the connecting relationship between the original cue and target stimuli. It has been argued that thought substitution and interference may still be able to play a role in SIF measured by IP tests via two stage models of recall (Tomlinson et al., 2009) or covert cuing (Camp et al., 2009). However, the current results indicate that these methods are insufficient as thought substitution instructions produce a very small and heterogeneous SIF effect as measured by IP tests. While the small number of studies makes it difficult to conclude that thought substitution produces no SIF effect in IP tests, based on the size of the difference relative to undifferentiated instructions and the level of confidence in the direction of the relationship, it is clear that they are associated with a large decrease in SIF.

In terms of the implication of this result, it has been suggested that IP testing is more analogous to real world recall than SP testing since people rarely establish a direct connection between two stimuli in order to prompt recall, instead typically being prompted by random related stimuli which they encounter during day-to-day activities (Anderson & Spellman, 1995). As such, this result may provide evidence that while thought substitution can produce the SIF effect, direct suppression is a more applicable mechanism in real world situations. More specifically, taken in combination with the previously discussed result that thought substitution is less effective when the substitutes are self-generated, it may be suggested that thought substitution would be unable to produce intentional forgetting in day-to-day life where recall more closely resembles IP testing than SP. Support for this idea is mixed, as historically clinical populations report ineffective thought control when employing thought replacement strategies (e.g., Freeston et al., 1995) while other studies suggest that distraction via thought substitution reduces intrusive thoughts (e.g., Lin & Wicker, 2007), although the ecological validity of these results is unclear. Additionally, there is insufficient research to determine if direct suppression is employed as an alternative strategy to thought substitution in real-world situations, making it difficult to draw firm conclusions regarding relative effectiveness. While further evidence is needed to strengthen a relationship between TNT and real-world SIF, it can be concluded based on the results of the present study that IP tests weaken the effectiveness of thought substitution instructions when producing SIF.

Repetitions and the SIF Effect

As SIF is considered to be an active and effort driven phenomenon (Anderson & Huddleston, 2020) it has been suggested that it has a cumulative effect on memory, meaning with increased opportunity to expend effort towards forgetting it either has a larger, additive effect or it becomes easier to suppress the target stimuli as participants become more adept at forgetting with practice. From this, it can be extrapolated that the SIF effect will increase linearly with the number of repetitions included in the TNT phase of the paradigm, a fact that has been supported across multiple studies (Anderson & Green, 2001; Hertel & Calcaterra, 2005; Lambert et al., 2010). Looking at more specific examples, Hulbert and Anderson (2018) demonstrated across two separate experiments that while the SIF effect was in fact present after a single repetition, it reliably increased in strength after a total of 16 repetitions. Similarly, Depue et al. (2006) found across two experiments that the SIF effect grew stronger with 10 repetitions relative to 5 repetitions.

To determine if this effect was consistent across the entire literature, it was hypothesized that the same positive relationship would extend over the entire range of repetitions included in TNT studies. This hypothesis was supported by the unconditionalized data which demonstrated a credible trend in the SP data and a non-credible but still positive trend in the IP data. The weaker relationship in studies using the IP tests may have been due to either the increased heterogeneity observed across all IP data or potentially due to the lower number of IP studies than SP studies. Regardless of the reason, the trend being positive does provide some support for the hypothesis, but the result may be considered less conclusive in IP data.

The hypothesis was not supported in the conditionalized data as the SP data demonstrated no effect of repetitions on SIF while the IP data demonstrated a very slight negative trend, both of which were non-credible. One possible explanation for this lack of effect is lower relative power in comparison to the models for the unconditionalized data as both SP and IP testing methods had fewer studies included in the conditionalized models. However, this explanation is less viable for the conditionalized data than for the unconditionalized IP data due simply to how less probable the effects are. As such, a more likely explanation for the result is that some part of the conditionalization process accounts for the variability previously explained by the number of repetitions. In an effort to determine the exact mechanism for this proposed interaction, a further examination of the conditionalized data found that only 2 of the studies reported over 95% no-think stimuli recall accuracy in the SP data and none did so in the IP data. This suggests that ceiling effects caused by conditionalization are unlikely to be the cause of the general decrease in heterogeneity in the conditionalized data relative to the unconditionalized data. Similarly, as shown in Figure 11 the distribution of repetitions included in the conditionalized literature is not noticeably more skewed than the unconditionalized literature, ranging from 1 to 18 repetitions. This suggests that the lack of effect is not related to a lack of variability in the number of repetitions.

A possible theoretical explanation for the results is that higher repetitions increased the SIF effect in the unconditionalized data by allowing for increased opportunity to recall items with weaker encoding that may have been incorrectly recalled in the learning phase. More specifically, being exposed to the cue stimuli a greater number of times during TNT may allow for improved recall of the associated targets which were not recalled in the initial test, making the targets more accessible and thus susceptible to SIF. As conditionalization removes the influence of items with weaker encoding and poorer recall in the learning phase, there would be no opportunity for repetition to have an effect, although this interpretation is solely hypothetical without further research. Regardless, it can be cautiously concluded that conditionalizing the data may reduce the impact of repetitions on the SIF effect although the exact mechanism is unclear. This may potentially indicate that the use of conditionalization could reduce the needed number of repetitions to observe a SIF effect in future experiments.

Stimuli Type and the SIF Effect

The stimuli type of the cue and target produced no credible trends in either the unconditionalized or conditionalized data when comparing words to non-word stimuli, which is in line with the hypothesis that words and non-word stimuli would produce similar SIF effect sizes. This result is somewhat surprising in the context of the overall memory literature, as historically different types of stimuli such as pictures and words have exhibited variability in their encoding and recall accuracy across different methods of testing (Park et al., 1983; Paivio & Csapo, 1971; Roediger, 1990). However, the TNT paradigm is thought to assess a basic cognitive process, namely active forgetting, which theoretically is used to indiscriminately supress the retrieval of specific memories regardless of characteristics such as stimuli type, complexity, or its associated neurological region (e.g., Anderson & Hulbert, 2021; Fawcett & Hulbert, 2020; Norby, 2018). This theory is supported by the simple fact that the SIF effect has been replicated within the literature across stimulus types including word pairs (Anderson & Green, 2001; Detre et al., 2013; Hanslmayr et al., 2009) as well as autobiographical memories (Noreen & MacLeod, 2013) and visual stimuli such as pictures and face-word pairs (Depue et al., 2006; Catarino et al., 2015). Additionally, there is neurological support as previous research has demonstrated that the TNT paradigm is capable of altering neural activity in regions such as the parahipocampal place area for spatial details, amygdala for emotional content, and fusiform cortex for faces (Depue et al., 2007, 2010; Gagnepain et al., 2014; Gagnepain et al., 2017; Mary et al., 2020).

As such, the goal of the present meta-analysis was not to determine whether SIF could affect memory of different types of stimuli but to instead determine the relative magnitude of the SIF effect when different types of cues and targets were used. This was partially motivated by a desire to determine if stimulus type played a role in studies which failed to replicate the SIF effect. However, the result indicated that within the existing literature stimuli type has no effect on SIF as participants were equally adept at forgetting pictures, memories and sounds as words. This may have practical implications concerning the SIF effect, as it provides evidence that SIF is applicable to stimuli regardless of complexity. More specifically, in terms of encoding and cognitive representations episodic memories are more complex than pictures which are in turn more complex than single words. TNT is thought to be an analogue for real world intentional forgetting (Anderson & Green, 2001), and the fact that it is able to replicate SIF in the above stimulus types (e.g., Catarino et al., 2015; Noreen & MacLeod, 2013) as well as words strengthens this claim. As a whole, this result provides limited theoretical insight due to the small number of studies but does clarify some of the boundary conditions of the SIF effect by demonstrating its capabilities to affect memory of different forms of stimuli.

Stimulus Emotional content and the SIF Effect

The emotional content of the cue and target stimuli was likewise unrelated to SIF in either the conditionalized or unconditionalized data, which is in line with the hypothesis that non-emotional and emotional stimuli would produce similar SIF effect sizes. However, there was a slight trend of emotional cues in the unconditionalized data producing a smaller SIF effect than non-emotional cues. This could be interpreted as emotional cues increasing the salience and depth of encoding of the paired target items relative to non-emotional cues, a finding which would be in line with the rationale and results of comparable studies examining clinical samples (Joorman et al., 2009; Marzi et al., 2014; Streb et al., 2016), but the result is inconclusive and based on a small number of studies. Additionally, in the only comparable meta-analysis to the present study, Stramaccia et al., (2021) found that in clinical samples across the entire literature emotional content had no effect on SIF, providing support for the present result in studies examining general populations. Overall, the present results suggest that stimulus emotional content does not relate to SIF, but due to the broad heterogeneity present in the observed effects it cannot be conclusively said if emotional content has an effect until further research is conducted on the topic.

In terms of theoretical perspectives, this result was not in line with what would be predicted based on the historical literature. Many cognitive theorists have proposed that emotional memories are better encoded than neutral comparisons due to mechanisms such as increased attention, elaboration, and encoding (Hamaan, 2001), resulting in increased accuracy for emotional items with the greatest effects being observed for negative items (Bradley & Mogg, 1994; Kern et al., 2005). Applied to the TNT literature, it logically follows that this may have an effect on both direct suppression and thought substitution, as memories which are more deeply encoded would likely be more difficult to suppress and the greater number of cognitive connections could lead to a higher potential for intrusions during TNT. Thought substitution may be more difficult to accomplish with emotional stimuli as research has demonstrated that emotional items are more salient during recall (Long et al., 2015; Strongman & Russel, 1986) and thus potential substitutes would need to be equally salient to achieve meaningful interference. In line with this logic, previous research has demonstrated that similar memory paradigms such as DF find increased resilience to intentional forgetting for emotional items relative to neutral items (Hall et al., 2021). Nonetheless, the present results indicate that the emotional content of the stimuli does not have a conclusive effect on SIF, suggesting that active forgetting is capable of overcoming these potential mitigating factors.

Returning to the basis of SIF, the TNT paradigm is meant to be an analogue of a naturally occurring cognitive process which effortfully forgets specific memories (Anderson & Green, 2001). Recent theorists have proposed that this and other similar cognitive processes serve the role of maintaining cognitive efficacy and health through the intentional forgetting of irrelevant or potentially harmful information (Fawcett & Huddleston, 2020). Following this logic, these processes would naturally be capable of influencing both neutral and emotional stimuli which would be encountered in day-to-day activities, providing theoretical support for the present result. Supporting this logic, some TNT studies (e.g., Depue et al., 2006) have found that emotional items are actually more susceptible to SIF than neutral items following the rationale that the items are more salient and thus more accessible, allowing for increased modification via suppression or interference processes. Alternatively, several studies have also demonstrated no difference between SIF for neutral and emotional items (e.g., Dieler et al., 2010; Murray et al., 2011), supporting that intentional forgetting is reliable regardless of emotional content. However, this line of thinking has historically been highly controversial within the literature as many researchers hold that suppression and active forgetting processes are ineffective or alternatively symptoms of psychopathology which can interfere with mental health and thus have no benefit (Becker et al., 1998; Erkskine et al., 2007; Wenzlaff & Wegner, 2000). Regardless of the theoretical arguments surrounding the subject, the present meta-analysis provides evidence that SIF is equally as effective for emotional stimuli as non-emotional, albeit in a limited number of studies.

Implications

Prior to determining the implications of the present study, the results must be discussed in relation to similar research. In the only comparable meta-analysis to the present thesis, Stramaccia et al. (2021) also found that there was a small aggregate SIF effect produced by TNT in the control populations of clinical samples. This provides additional support for our main finding of a reliable small SIF effect in general populations over a much larger sample of studies. In terms of moderators, the other study matched our results of emotional content having no effect on SIF and more specific instructions being associated with larger SIF effects while also contradicting our results concerning stimuli type by finding that pictures produced a larger effect than other stimuli and that repetitions had no effect. The latter two findings as well as the overall larger SIF effect sizes reported throughout the study may be explained by the chosen sample, namely that they selectively examined SIF in clinical control populations. More specifically, Stramaccia and colleagues primarily examined control populations who had been screened for clinical disorders, and such populations may represent individuals who are more adept at controlling their thoughts than the average person (e.g., Warren et al., 2021). As such, these participants may exhibit greater than average SIF, a fact which would explain the contradicting results for stimuli type and repetitions. Further, the study's analysis of instruction as a moderator may overexaggerate the effect in the general population as exemplified by their finding that direct suppression results in a moderate SIF effect size compared to the small effect size observed in the present thesis. Overall, while the other meta-analysis provides valuable information concerning SIF in clinical populations, I would argue that the present meta-analysis provides a more accurate representation of SIF in the general population by excluding clinical control samples and including a much larger sample of studies (82 vs 25).

The results of the present thesis have implications not only on the various theoretical perspectives described above but also on real world applications of memory research. The most notable application currently explored in the TNT literature would be the impact of intentional forgetting in clinical settings ranging from mood disorders to eating disorders and substance abuse disorders. Specifically, TNT and SIF have been examined in several studies in the context of intrusive thoughts as a result of anxiety (e.g., Benoit et al., 2016), trauma (e.g., Streb et al., 2016), and PTSD (e.g., Catarino et al., 2016) with the reliable result that participants with clinical disorders or those scoring high on clinical measures fail to demonstrate SIF. Further, multiple studies have demonstrated that the SIF effect is weaker for individuals with severe self-reported ruminative tendencies (Fawcett et al., 2015, Joorman et al., 2005). These results are consistent with those of Stramaccia et al. (2021), suggesting that this trend extends across the entire clinical literature which has to date been examined in the context of TNT. The results of the present thesis, along with those of Strammaccia et al, demonstrated that the typical individual is capable of intentional forgetting across a variety of conditions and using either direct suppression or thought substitution while clinical populations cannot. In terms of the implications of this result on clinical populations, it provides support for the idea that the development of pathological intrusive thoughts is somehow related to a deficit in executive control processes surrounding memory such as SIF, although it is unknown based on the existing literature if this relationship is causal. More specifically, it is currently unclear if deficits in thought control such as SIF play a role in the development of clinical disorders or alternatively if said deficits are instead a symptom of clinical disorders. Future research should examine if TNT and other intentional forgetting paradigms may be able to identify neurological and cognitive factors associated with the presence or development of clinical disorders.

Besides clinical implications, the most notable practical application currently explored in the TNT literature is the impact of intentional forgetting on criminal testimony in law enforcement settings (e.g., Bergström et al., 2013; Hu et al., 2015). Specifically, there is the concern that individuals who commit crimes may be capable of intentionally forgetting specific details which may interfere in establishing their guilt or innocence. Alternatively, the SIF effect may support the idea that witnesses are capable of self-altering their memory of events they consider unpleasant or distressing, bringing into question the reliability of eyewitness testimony which can commonly be thought of as highly reliable (Magnussen et al., 2009). As our results imply that SIF can have a reliable effect not only on minor details but also on more complex stimuli such as episodic memories (e.g., Noreen & MacLeod, 2013), they provide support for continued examination of how individuals may modify their memories in such situations. More generally, the present results support the use of the TNT paradigm as an analogue of intentional forgetting in efforts to demonstrate the real-world implications of SIF. Additionally, our results provide support for perspectives which suggest that intentional forgetting may be a viable method for improving the efficacy of cognition, maintaining cognitive flexibility by allowing for the removal of outdated information and the removal of unwanted or unneeded memories (Fawcett & Huddleston, 2020). Thus, it may be a viable area of research to examine if TNT and the strategies implemented within the paradigm, such as encouraging direct suppression or thought substitution, may be able to help individuals with deficits in these areas of cognition as well as further our understanding of SIF.

Limitations

The major limitation of the present study was the inability to draw firmer conclusions regarding some of the studied moderators due to a limited number of available studies. More specifically, the analyses of repetitions in the conditionalized data as well as stimulus type and emotional content across the entire literature may have been overly influenced by having much fewer studies than the other, more highly powered moderator models. Looking at a specific example, stimulus emotional content tended to be more thoroughly studied in clinical populations (Joorman et al., 2009; Marzi et al., 2014; Streb et al., 2016) relative to the general sample examined in the current study. This potentially indicates an area of need for further research in general populations but presently represents a limitation in the ability to define the effect of emotional content more conclusively. As such, while several of the moderators can be conclusively said to have or not have an effect on SIF, the results of the present study should be reanalysed as more research continues to become available to maximize its informative qualities on the theoretical and practical literature.

Future Directions

Overall, our findings demonstrated that there is a reliable, albeit small SIF effect produced by the TNT paradigm regardless of chosen testing and conditionalization method. Despite this finding, there are still a number of studies which have failed to replicate the SIF effect (e.g., Bulevich et al., 2006). One possible explanation for this inconsistency is that some studies may be underpowered and thus unable to detect the effect. Based on the aggregate effects identified through our models, we calculated through the use of the program G*POWER (Erdfelder et al., 1996) the needed sample size to find the SIF effect in an unconditionalized study with a minimum power of 0.8 is 75 and 109 for SP and IP tests, respectively. For the studies included in the present meta-analysis the average sample size for SP and IP tests was 33 and 36, respectively, which indicates that the typical SP study has a power of 0.49 while the typical IP study has a power of only 0.38. This issue is likely due to the time intensive nature of the TNT paradigm which limits both the available participants as well as the viability of conducting a sufficiently powered study. However, based on the results of the present thesis there are methodological decisions which may influence the size of the SIF effect and thus lower the required participants, making it easier to conduct accurate and reliable TNT studies.

The most consistent methodological recommendation is that all studies should at least contain a comparison between the unconditionalized and conditionalized results. Based on the results of the present meta-analysis, conditionalizing the results lowers the required sample size to 50 for both SP and IP studies, lowering the barrier to entry for finding the SIF effect. Alternatively, if future research prefers to use only unconditionalized data, researchers should attempt to increase the number of repetitions during the TNT phase, although this will likely increase the needed time for the study. There are also likely diminishing returns past a certain number of repetitions where the memory being targeted cannot be further suppressed (i.e., successful recall is at 0%). Additionally, increasing the number of repetitions also increases the delay between learning and testing of baseline items, potentially reducing baseline recall accuracy at higher numbers of repetitions and thus reducing the size of the SIF effect observed when comparing no-think accuracy to baseline accuracy. As the present analysis examines the relationship solely in a linear fashion, further research is needed to examine potential nonlinear relationships involved and establish the number of repetitions with the greatest impact on SIF. As a compromise based on the current analyses it is recommended that at least 16 repetitions be included, the most commonly used amount across the literature. Finally, future research using SP tests should give more specific instructions (i.e., direct suppression or thought substitution) to participants, unless specifically examining the effect of undifferentiated instructions, while studies using IP tests should avoid giving thought substitution instructions. All of these methods have been demonstrated across the literature examined in the present meta-analysis to maximize the resulting SIF effect and thus allow for more thorough theoretical examinations of SIF.

Conclusions

In conclusion, the present meta-analysis provided a description of the typical SIF effect produced by the TNT paradigm and resolved the ambiguity concerning the effect by demonstrating that the paradigm can produce a reliable effect dependent on the methodological choices made in each study. While many of the moderators such as repetitions, instruction, and conditionalization were able to demonstrate conclusive effects on SIF, others such as stimulus emotional content and type were somewhat less conclusive and should continue to be evaluated as more research on the subject is conducted. Overall, the moderator models demonstrated that the SIF effect was (a) larger for conditionalized data than for unconditionalized data; (b) slightly smaller and more heterogenous for studies using the IP testing method relative to SP; (c) larger when participants were given direct suppression or thought substitution instructions instead of less specific instructions for SP testing methods but not IP; (d) larger when participants were given thought substitutes to use relative to when they generated their own substitutes; (e) larger at greater numbers of repetitions during the think/no-think phase in unconditionalized studies but not conditionalized studies; and, (f) unaffected by stimulus type or emotional content. Overall, the present results indicate that SIF is a reliable effect and researchers should continue to explore its theoretical and practical implications in an effort to deepen our understanding of memory and overall cognition.

References

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, *24*(3), 437-448. https://doi.org/10.1007/s10648-012-9210-2

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: the role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General*, *130*(3), 544. https://doi.org/10.1037/0096-3445.130.3.544

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063. https://doi.org/10.1037/0278-7393.20.5.1063

*Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366-369. https://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*(5), 1141. https://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*(6), 279-292. https://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). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-1195-6_3

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*, *72*. https://doi.org/10.1146/annurev-psych-072720-094140

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–102). American Psychological Association. https://doi.org/10.1037/11587-005

Anderson, M. C., & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science*, *18*(4), 189-194. https://doi.org/10.1111/j.1467-8721.2009.01634.x

*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*(5655), 232-235. https://doi.org/10.1126/science.1089504

*Anderson, M. C., Reinholz, J., Kuhl, B. A., & Mayr, U. (2011). Intentional suppression of unwanted memories grows more difficult as we age. *Psychology and aging*, *26*(2), 397. https://doi.org/10.1037/a0022505

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological review*, *102*(1), 68. https://doi.org/10.1037/0033-295X.102.1.68

Aslan, A., Bäuml, K. H., & Grundgeiger, T. (2007). The role of inhibitory processes in part-list cuing. *Journal of experimental psychology: Learning, Memory, and Cognition*, *33*(2), 335. https://doi.org/10.1037/0278-7393.33.2.335

*Attuquayefio, T., Stevenson, R. J., Boakes, R. A., Oaten, M. J., Yeomans, M. R., Mahmut, M., & Francis, H. M. (2016). A high-fat high-sugar diet predicts poorer hippocampal-related memory and a reduced ability to suppress wanting under satiety. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*(4), 415. https://doi.org/10.1037/xan0000118

Bailey, K., & Chapman, P. (2012). When can we choose to forget? An ERP study into item-method directed forgetting of emotional words. *Brain and Cognition*, *78*(2), 133-147. https://doi.org/10.1016/j.bandc.2011.11.004

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603. https://doi.org/10.1037/0278-7393.19.3.603

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257-278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Becker, E. S., Rinck, M., Roth, W. T., & Margraf, J. (1998). Don't worry and beware of white bears: Thought suppression in anxiety patients. *Journal of anxiety Disorders*, *12*(1), 39-55. https://doi.org/10.1016/S0887-6185(97)00048-0

*Bell, T. A. (2005). *Individual differences in memory inhibition* (Unpublished doctoral dissertation). University of Oregon, Eugene, United States.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic bulletin & review*, *13*(5), 831-836. https://doi.org/10.3758/BF03194005

*Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*, *76*(2), 450-460. https://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*, *113*(52), E8492-E8501. https://doi.org/10.1073/pnas.1606604114

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

*Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*(4), 726-737. https://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. https://doi.org/10.1016/j.brainres.2007.03.014

Bjork, R. A. (1972). Theoretical explanations of directed forgetting. *Coding process in human memory, Washington, DC: Winston & Sons*.

Bjork, R. A., & Woodward, A. E. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology*, *99*(1), 22. https://doi.org/10.1037/h0034757

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). Lawrence Erlbaum Associates, Inc.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods*, *8*(1), 5-18. https://doi.org/10.1002/jrsm.1230

Bradley, B. P., & Mogg, K. (1994). Mood and personality in recall of positive and negative information. *Behaviour research and therapy*, *32*(1), 137-141. https://doi.org/10.1016/0005-7967(94)90095-7

Breuer, J., & Freud, S. (1955). On the psychical mechanism of hysterical phenomena: Preliminary communication from studies on hysteria. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume II (1893-1895): Studies on Hysteria* (pp. 1-17).

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology*, *10*(1), 12-21. https://doi.org/10.1080/17470215808416249

*Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & cognition*, *34*(8), 1569-1577. https://doi.org/10.3758/BF03195920

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1-28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395-411. https://arxiv.org/pdf/1705.11123.pdf

Butler, A. J., & James, K. H. (2010). The neural correlates of attempting to suppress negative versus neutral memories. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 182-194. https://doi.org/10.3758/CABN.10.2.182

*Cano, M. E., & Knight, R. T. (2016). Behavioral and EEG evidence for auditory memory suppression. *Frontiers in human neuroscience*, *10*, 133. https://doi.org/10.3389/fnhum.2016.00133

Camp, G., Pecher, D., & Schmidt, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review*, *12*(3), 490-494. https://doi.org/10.3758/BF03193793

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633-642. https://doi.org/10.3758/BF03202713

*Castiglione, A., Wagner, J., Anderson, M., & Aron, A. R. (2019). Preventing a thought from coming to mind elicits increased right frontal beta just as stopping action does. *Cerebral Cortex*, *29*(5), 2160-2172. https://doi.org/10.1093/cercor/bhz017

*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*(5), 604-616. https://doi.org/10.1177/0956797615569889

Camp, G., Pecher, D., Schmidt, H. G., & Zeelenberg, R. (2009). Are independent probes truly independent?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 934. https://doi.org/10.1037/a0015536

*Chen, C., Liu, C., Huang, R., Cheng, D., Wu, H., Xu, P., ... & Luo, Y. J. (2012). Suppression of aversive memories associates with changes in early and late stages of neurocognitive processing. *Neuropsychologia*, *50*(12), 2839-2848. https://doi.org/10.1016/j.neuropsychologia.2012.08.004

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1403. https://doi.org/10.1037/0278-7393.25.6.1403

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of verbal learning and verbal behavior*, *12*(6), 599-607. https://doi.org/10.1016/S0022-5371(73)80039-8

*Davidson, P., Hellerstedt, R., Jönsson, P., & Johansson, M. (2020). Suppression-induced forgetting diminishes following a delay of either sleep or wake. *Journal of Cognitive Psychology*, *32*(1), 4-26. https://doi.org/10.1080/20445911.2019.1705311

*De Vito, D., & Fenske, M. J. (2017). Suppressing memories of words and familiar objects results in their affective devaluation: Evidence from Think/No-think tasks. *Cognition*, *162*, 1-11. https://doi.org/10.1016/j.cognition.2017.01.020

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language*, *52*(1), 120-130. https://doi.org/10.1016/j.jml.2004.09.002

*del Prete, F., Hanczakowski, M., Bajo, M. T., & Mazzoni, G. (2015). Inhibitory effects of thought substitution in the think/no-think task: Evidence from independent cues. *Memory*, *23*(4), 507-517. https://doi.org/10.1080/09658211.2014.907429

DePrince, A. P., & Freyd, J. J. (2004). Forgetting trauma stimuli. *Psychological Science*, *15*(7), 488-492. https://doi.org/10.1111/j.0956-7976.2004.00706.x

*Depue, B. E., Banich, M. T., & Curran, T. (2006). Suppression of emotional and nonemotional content in memory: Effects of repetition on cognitive control. *Psychological science*, *17*(5), 441-447. https://doi.org/10.1111/j.1467-9280.2006.01725.x

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*(13), 3909-3917. https://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*(5835), 215-219. https://doi.org/10.1126/science.1139560

*Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and neural oscillations during volitional suppression of memory retrieval. *Journal of cognitive neuroscience*, *25*(10), 1624-1633. https://doi.org/10.1162/jocn_a_00418

*Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., & Banich, M. T. (2016). The organization of right prefrontal networks reveals common mechanisms of inhibitory regulation across cognitive, emotional, and motor processes. *Cerebral Cortex*, *26*(4), 1634-1646. https://doi.org/10.1093/cercor/bhu324

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*(12), 2371-2388. https://doi.org/10.1016/j.neuropsychologia.2013.02.017

*Dieler, A. C. (2011). Investigation of variables influencing cognitive inhibition: from the behavioral to the molecular level (Unpublished doctoral dissertation). Julius Maximilian University of Würzburg, Wurzburg, Germany.

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*(3), 184-193. https://doi.org/10.1080/09658211.2013.774420

*Dieler, A. C., Plichta, M. M., Dresler, T., & Fallgatter, A. J. (2010). Suppression of emotional words in the Think/No-Think paradigm investigated with functional near-infrared spectroscopy. *International Journal of Psychophysiology*, *78*(2), 129-135. https://doi.org/10.1016/j.ijpsycho.2010.06.358

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28*(1), 1–11. https://doi.org/10.3758/BF03203630

Erskine, J. A., Kvavilashvili, L., & Kornbrot, D. E. (2007). The predictors of thought suppression in young and old adults: Effects of rumination, anxiety, and other variables. *Personality and Individual Differences*, *42*(6), 1047-1057. https://doi.org/10.1016/j.paid.2006.09.016

*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. https://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 and Cognition*. https://doi.org/10.1016/j.jarmac.2019.11.002

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition*, *36*(6), 1168-1181. https://doi.org/10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition*, *38*(6), 797-808. https://doi.org/10.3758/MC.38.6.797

*Fischer, S., Diekelmann, S., & Born, J. A. N. (2011). Sleep's role in the processing of unwanted memories. *Journal of sleep research*, *20*(2), 267-274. https://doi.org/10.1111/j.1365-2869.2010.00881.x

Ferreira, C. S., Marful, A., Staudigl, T., Bajo, T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *Journal of Cognitive Neuroscience*, *26*(4), 777-791. https://doi.org/10.1162/jocn_a_00523

Freeston, M. H., Ladouceur, R., Provencher, M., & Blais, F. (1995). Strategies used with intrusive thoughts: Context, appraisal, mood, and efficacy. *Journal of anxiety disorders*, *9*(3), 201-215. https://doi.org/10.1016/0887-6185(95)00002-6

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*, *111*(13), E1310-E1319. https://doi.org/10.1073/pnas.1311468111

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. *Journal of Neuroscience*, *37*(27), 6423-6441. https://doi.org/10.1523/JNEUROSCI.2732-16.2017

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (Vol. 1). New York, NY, USA: Cambridge University Press.

*Gilchrist, A. L. (2005). MEMORY SUPPRESSION IN THE THINK/NO-THINK PARADIGM: THE ROLE OF STRATEGY [Unpublished honors thesis]. Florida State University, Tallahassee, United States

*Gillie, B. L., Vasey, M. W., & Thayer, J. F. (2014). Heart rate variability predicts control over memory retrieval. *Psychological Science*, *25*(2), 458-465. https://doi.org/10.1177/0956797613508789

Gómez-Ariza, C. J., Del Prete, F., Prieto del Val, L., Valle, T., Bajo, M. T., & Fernandez, A. (2017). Memory inhibition as a critical factor preventing creative problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 986. https://doi.org/10.1037/xlm0000348

Guo, Y., Schmitz, T. W., Mur, M., Ferreira, C. S., & Anderson, M. C. (2018). A supramodal role of the basal ganglia in memory and motor inhibition: Meta-analytic evidence. *Neuropsychologia*, *108*, 117-134. https://doi.org/10.1016/j.neuropsychologia.2017.11.033

Hall, K. J., Fawcett, E. J., Hourihan, K. L., & Fawcett, J. M. (2021). Emotional memories are (usually) harder to forget: A meta-analysis of the item-method directed forgetting literature. *Psychonomic Bulletin & Review*, 1-14. https://doi.org/10.3758/s13423-021-01914-z

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences*, *5*(9), 394-400. https://doi.org/10.1016/S1364-6613(00)01707-1

Hayashino, Y., Noguchi, Y., & Fukui, T. (2005). Systematic evaluation and comparison of statistical tests for publication bias. *Journal of epidemiology*, *15*(6), 235-243. https://doi.org/10.2188/jea.15.235

Hanczakowski, M., Pasek, T., & Zawadzka, K. (2012). Context-dependent impairment of recollection in list-method directed forgetting. *Memory*, *20*(7), 758-770. https://doi.org/10.1080/09658211.2012.702774

*Hanslmayr, S., Leipold, P., & Bäuml, K. H. (2010). Anticipation boosts forgetting of voluntarily suppressed memories. *Memory*, *18*(3), 252-257. https://doi.org/10.1080/09658210903476548

*Hanslmayr, S., Leipold, P., Pastötter, B., & Bäuml, K. H. (2009). Anticipatory signatures of voluntary memory suppression. *Journal of Neuroscience*, *29*(9), 2742-2747. https://doi.org/10.1523/JNEUROSCI.4703-08.2009

Heijltjes, A., van Gog, T., Leppink, J., & Paas, F. (2015). Unraveling the effects of critical thinking instructions, practice, and self-explanation on students' reasoning performance. *Instructional Science*, *43*(4), 487-506. https://doi.org/10.1007/s11251-015-9347-8

*Hellerstedt, R., Johansson, M., & Anderson, M. C. (2016). Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia*, *89*, 510-523. https://doi.org/10.1016/j.neuropsychologia.2016.07.008

*Herbert, C., & Sütterlin, S. (2012). Do not respond! Doing the think/no-think and go/no-go tasks concurrently leads to memory impairment of unpleasant items during later recall. *Frontiers in psychology*, *3*, 269. https://doi.org/10.3389/fpsyg.2012.00269

*Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, *12*(3), 484-489. https://doi.org/10.3758/BF03193792

Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological Science*, *14*(6), 573-578. https://doi.org/10.1046/j.0956-7976.2003.psci_1467.x

*Hertel, P. T., & Hayes, J. A. (2015). Distracted by cues for suppressed memories. *Psychological science*, *26*(6), 775-783. https://doi.org/10.1177/0956797615570711

Hertel, P., & McDaniel, L. (2010). The suppressive power of positive thinking: Aiding suppression-induced forgetting in repressive coping. *Cognition and Emotion*, *24*(7), 1239-1249. https://doi.org/10.1080/02699930903172377

*Hotta, C., & Kawaguchi, J. (2009). Self-initiated use of thought substitution can lead to long term forgetting. *Psychologia*, *52*(1), 41-49. https://doi.org/10.2117/psysoc.2009.41

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, *26*(7), 1098-1106. https://doi.org/10.1177/0956797615575734

Hu, X., Bergström, Z. M., Gagnepain, P., & Anderson, M. C. (2017). Suppressing unwanted memories reduces their unintended influences. *Current Directions in Psychological Science*, *26*(2), 197-206. https://doi.org/10.1177/0963721417689881

Huber, D. E., Tomlinson, T. D., Rieth, C. A., & Davelaar, E. J. (2010). Reply to Bäuml and Hanslmayr: Adding or subtracting memories? The neural correlates of learned interference vs. memory inhibition. *Proceedings of the National Academy of Sciences*, *107*(2), E4-E4. https://doi.org/10.1073/pnas.0912815107

Huber, D. E., Tomlinson, T. D., & Jang, Y. (2015). The search of associative memory with recovery interference (SAM-RI) memory model and its application to retrieval practice paradigms. In *Cognitive Modeling in Perception and Memory* (pp. 91-108). Psychology Press.

Huddleston, E., & Anderson, M. C. (2012). Reassessing critiques of the independent probe method for studying inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1408. https://doi.org/10.1037/a0027092

*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*(12), 1931. http://doi.org/10.1037/xge0000461

Hulbert, J. C., Henson, R. N., & Anderson, M. C. (2016). Inducing amnesia through systemic suppression. *Nature Communications*, *7*(1), 1-9. https://doi.org/10.1038/ncomms11003

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7), e010247. http://dx.doi.org/10.1136/bmjopen-2015-010247

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and clinical neurophysiology*, *82*(6), 477-482. https://doi.org/10.1016/0013-4694(92)90054-L

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science*, *15*(7), 448-453. https://doi.org/10.1111/j.0956-7976.2004.00700.x

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*(4), 640. https://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*(1), 34. https://doi.org/10.1037/a0013794

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *science*, *319*(5865), 966-968. https://doi.org/10.1126/science.1152408

Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P., & Ridderinkhof, K. R. (2004). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology*, *41*(1), 9-20. https://doi.org/10.1046/j.1469-8986.2003.00127.x

Kern, R. P., Libkuman, T. M., Otani, H., & Holmes, K. (2005). Emotional stimuli, divided attention, and memory. *Emotion*, *5*(4), 408. https://doi.org/10.1037/1528-3542.5.4.408

*Klein, K. (2014). *Suppression of free associations in the think/no-think paradigm* (Unpublished doctoral dissertation). Colorado State University, Fort Collins, United States

*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*(4), 1443. https://doi.org/10.1037/a0036518

*Lambert, A. J., Good, K. S., & Kirk, I. J. (2010). Testing the repression hypothesis: Effects of emotional valence on memory suppression in the think–no think task. *Consciousness and cognition*, *19*(1), 281-293. https://doi.org/10.1016/j.concog.2009.09.004

Lee, Y. S. (2018). Withdrawal of spatial overt attention following intentional forgetting: evidence from eye movements. *Memory*, *26*(4), 503-513. https://doi.org/10.1080/09658211.2017.1378360

*Lee, Y. S., Lee, H. M., & Tsai, S. H. (2007). Effects of post-cue interval on intentional forgetting. *British Journal of Psychology*, *98*(2), 257-272. https://doi.org/10.1348/000712606X120410

*Legrand, N., Etard, O., Vandevelde, A., Pierre, M., Viader, F., Clochon, P., ... & Gagnepain, P. (2018). Does the heart forget? Modulation of cardiac activity induced by inhibitory control over emotional memories. *bioRxiv*, 376954. https://doi.org/10.1101/376954

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 970. https://doi.org/10.1037/a0015728

*LeMoult, J., Hertel, P. T., & Joormann, J. (2010). Training the forgetting of negative words: The role of direct suppression and the relation to stress reactivity. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *24*(3), 365-375. https://doi.org/10.1002/acp.1682

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of personality and social psychology*, *71*(6), 1092. https://doi.org/10.1037/0022-3514.71.6.1092

Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta psychologica*, *127*(3), 623-635. https://doi.org/10.1016/j.actpsy.2007.12.004

*Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *Journal of Neuroscience*, *32*(47), 16785-16794. https://doi.org/10.1523/JNEUROSCI.2640-12.2012

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, *18*(1), 29-34. https://doi.org/10.1111/j.1467-9280.2007.01844.x

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory*, *19*(4), 346-359. https://doi.org/10.1080/09658211.2011.569725

Lin, Y. J., & Wicker, F. W. (2007). A comparison of the effects of thought suppression, distraction and concentration. *Behaviour Research and Therapy*, *45*(12), 2924-2937. https://doi.org/10.1016/j.brat.2007.08.004

*Liu, Y., Lin, W., Liu, C., Luo, Y., Wu, J., Bayley, P. J., & Qin, S. (2016). Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. *Nature communications*, *7*(1), 1-12. https://doi.org/10.1038/ncomms13375

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic bulletin & review*, *22*(5), 1328-1333. https://doi.org/10.3758/s13423-014-0791-2

López-Caneda, E., Crego, A., Campos, A. D., González-Villar, A., & Sampaio, A. (2019). The Think/No-Think Alcohol Task: A New Paradigm for Assessing Memory Suppression in Alcohol-Related Contexts. *Alcoholism: Clinical and Experimental Research*, *43*(1), 36-47. https://doi.org/10.1111/acer.13916

Ludowig, E., Möller, J., Bien, C. G., Münte, T. F., Elger, C. E., & Rosburg, T. (2010). Active suppression in the mediotemporal lobe during directed forgetting. *Neurobiology of Learning and Memory*, *93*(3), 352-361. https://doi.org/10.1016/j.nlm.2009.12.001

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(3), 271. https://doi.org/10.1037/0278-7393.1.3.271

Golding, J.M., & Macleod, C.M. (1998). Intentional forgetting: Interdisciplinary approaches. *Canadian Journal of Experimental Psychology, 52*, 160. https://doi.org/10.4324/9780203774434

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123-129. https://doi.org/10.3758/BF03210819

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3–23). American Psychological Association. https://doi.org/10.1037/11587-001

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671. https://doi.org/10.1037/a0018785

MacLeod, M. D., & Saunders, J. (2005). The role of inhibitory control in the production of misinformation effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 964. https://doi.org/10.1037/0278-7393.31.5.964

Magnussen, S., Melinder, A., Stridbeck, U., & Raja, A. Q. (2010). Beliefs about factors affecting the reliability of eyewitness testimony: A comparison of judges, jurors and the general public. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *24*(1), 122-133. https://doi.org/10.1002/acp.1550

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., ... & Peschanski, D. (2020). Resilience after trauma: The role of memory suppression. *Science*, *367*(6479). https://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. https://doi.org/10.3389/fpsyg.2013.01001

*Meier, B., König, A., Parak, S., & Henke, K. (2011). Suppressed, but not forgotten. *Swiss Journal of Psychology*. https://doi.org/10.1024/1421-0185/a000033

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

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *The American Journal of Psychology*, *53*(2), 173-203. https://doi.org/10.2307/1417415

*Murray, B. D., Anderson, M. C., & Kensinger, E. A. (2015). Older adults can suppress unwanted memories when given an appropriate strategy. *Psychology and Aging*, *30*(1), 9. http://doi.org/10.1037/a0038611

Murray, B. D., Muscatell, K. A., & Kensinger, E. A. (2011). Effects of emotion and age on performance during a think/no-think memory task. *Psychology and aging*, *26*(4), 940. https://doi.org/10.1037/a0023214

*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. https://doi.org/10.3389/fpsyg.2014.01396

Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral Cortex*, *20*(11), 2760-2770. https://doi.org/10.1093/cercor/bhq021

Nørby, S. (2018). Forgetting and emotion regulation in mental health, anxiety and depression. *Memory*, *26*(3), 342-363. https://doi.org/10.1080/09658211.2017.1346130

*Nørby, S., Lange, M., & Larsen, A. (2010). Forgetting to forget: On the duration of voluntary suppression of neutral and emotional memories. *Acta psychologica*, *133*(1), 73-80. https://doi.org/10.1016/j.actpsy.2009.10.002

*Noreen, S., & de Fockert, J. W. (2017). The role of cognitive load in intentional forgetting using the think/no-think task. *Experimental Psychology*. https://doi.org/10.1027/1618-3169/a000347

*Noreen, S., & MacLeod, M. D. (2013). It's all in the detail: Intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 375. https://doi.org/10.1037/a0028888

*Noreen, S., O'Connor, A. R., & MacLeod, M. D. (2016). Neural correlates of direct and indirect suppression of autobiographical memories. *Frontiers in Psychology*, *7*, 379. https://doi.org/10.3389/fpsyg.2016.00379

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

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

Nowicka, A., Marchewka, A., Jednorog, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: an fMRI study of directed forgetting. *Cerebral Cortex*, *21*(3), 539-549. https://doi.org/10.1093/cercor/bhq117

Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus, E. F. (2019). The return of the repressed: The persistent and problematic claims of long-forgotten trauma. *Perspectives on psychological science*, *14*(6), 1072-1095. https://doi.org/10.1177/1745691619862306

Paivio, A., & Csapo, K. (1971). Short-term sequential memory for pictures and words. *Psychonomic Science*, *24*(2), 50-51. https://doi.org/10.3758/BF03337887

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of gerontology*, *38*(5), 582-588. https://doi.org/10.1093/geronj/38.5.582

*Paz-Alonso, P. M., Bunge, S. A., Anderson, M. C., & Ghetti, S. (2013). Strength of coupling within a mnemonic control network differentiates those who can and cannot suppress memory retrieval. *Journal of Neuroscience*, *33*(11), 5017-5026. https://doi.org/10.1523/JNEUROSCI.3459-12.2013

*Paz-Alonso, P. M., Ghetti, S., Matlen, B. J., Anderson, M. C., & Bunge, S. A. (2009). Memory suppression is an active process that improves over childhood. *Frontiers in Human Neuroscience*, *3*, 24. https://doi.org/10.3389/neuro.09.024.2009

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *64*(1), 41. https://doi.org/10.1037/a0016569

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raaijmakers, J. G., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of memory and language*, *68*(2), 98-122.forgetting. *Cerebral Cortex*, *21*(3), 539-549. https://doi.org/10.1016/j.jml.2012.10.002

Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In *Psychology of learning and motivation* (Vol. 14, pp. 207-262). Academic Press. https://doi.org/10.1016/S0079-7421(08)60162-0

*Racsmány, M., Conway, M. A., Keresztes, A., & Krajcsi, A. (2012). Inhibition and interference in the think/no-think task. *Memory & cognition*, *40*(2), 168-176. https://doi.org/10.3758/s13421-011-0144-6

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American psychologist*, *45*(9), 1043. https://doi.org/10.1037//0003-066x.45.9.1043

Rizio, A. A., & Dennis, N. A. (2013). The neural correlates of cognitive control: successful remembering and intentional forgetting. *Journal of Cognitive Neuroscience*, *25*(2), 297-312. https://doi.org/10.1162/jocn_a_00310

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, *15*(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

*Ryckman, N. A., Addis, D. R., Latham, A. J., & Lambert, A. J. (2018). Forget about the future: effects of thought suppression on memory for imaginary emotional episodes. *Cognition and Emotion*, *32*(1), 200-206. https://doi.org/10.1080/02699931.2016.1276049

*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*(1), 77-93. https://doi.org/10.3758/s13415-016-0464-x

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. *Psychology of learning and motivation*, *59*, 131-189. https://doi.org/10.1016/B978-0-12-407187-2.00004-6

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*(4), 464-476. https://doi.org/10.3758/MC.37.4.464

*Sakaki, M., Kuhbandner, C., Mather, M., & Pekrun, R. (2014). Memory suppression can help people "unlearn" behavioral responses—but only for nonemotional memories. *Psychonomic bulletin & review*, *21*(1), 136-141. https://doi.org/10.3758/s13423-013-0480-6

*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: JINS*, *13*(2), 277. https://doi.org/10.1017/S1355617707070300

Salthouse, T. A. (2009). When does age-related cognitive decline begin?. *Neurobiology of aging*, *30*(4), 507-514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

*Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescot, A. P., & Anderson, M. C. (2017). Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nature communications*, *8*(1), 1-12. https://doi.org/10.1038/s41467-017-00956-z

Scholz, S., & Dutke, S. (2019). Investigating intentional forgetting using pupillometry: no evidence for the ongoing allocation of cognitive resources during forgetting. *Journal of Cognitive Psychology*, *31*(4), 416-437. https://doi.org/10.1080/20445911.2019.1622548

Seli, P., Schacter, D. L., Risko, E. F., & Smilek, D. (2019). Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychological Research*, *83*(5), 1057-1069. https://doi.org/10.1007/s00426-017-0914-2

Shi, X., Nie, C., Shi, S., Wang, T., Yang, H., Zhou, Y., & Song, X. (2017). Effect comparison between Egger's test and Begg's test in publication bias diagnosis in meta-analyses: evidence from a pilot survey. *Int J Res Stud Biosci*, *5*(5), 14-20. http://dx.doi.org/10.20431/2349-0365.0505003

Spitzer, B. (2014). Finding retrieval-induced forgetting in recognition tests: a case for baseline memory strength. *Frontiers in psychology*, *5*, 1102. https://doi.org/10.3389/fpsyg.2014.01102

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of clinical epidemiology*, *53*(11), 1119-1129. https://doi.org/10.1016/S0895-4356(00)00242-0

Stramaccia, D., Rischer, K. M., Fawcett, J. M., & Benoit, R. G. (2021). Memory suppression and its deficiency in psychological disorders: A focused meta-analysis. https://doi.org/10.1037/xge0000971

*Streb, M., Mecklinger, A., Anderson, M. C., Lass-Hennemann, J., & Michael, T. (2016). Memory control ability modulates intrusive memories after analogue trauma. *Journal of affective disorders*, *192*, 134-142. https://doi.org/10.1016/j.jad.2015.12.032

Strongman, K. T., & Russell, P. N. (1986). Salience of emotion in recall. *Bulletin of the Psychonomic Society*, *24*(1), 25-27. https://doi.org/10.3758/BF03330493

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435–452). Russell Sage Foundation.

*Tramoni, E., Aubert-Khalfa, S., Guye, M., Ranjeva, J. P., Felician, O., & Ceccaldi, M. (2009). Hypo-retrieval and hyper-suppression mechanisms in functional amnesia. *Neuropsychologia*, *47*(3), 611-624. https://doi.org/10.1016/j.neuropsychologia.2008.11.012

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*(37), 15588-15593. https://doi.org/10.1073/pnas.0813370106

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of verbal learning and verbal behavior*, *6*(2), 175-184. https://doi.org/10.1016/S0022-5371(67)80092-6

Underwood, B. J. (1957). Interference and forgetting. *Psychological review*, *64*(1), 49. https://doi.org/10.1037/h0044616

Valle, T. M., Gómez-Ariza, C. J., & Bajo, M. T. (2019). Inhibitory control during selective retrieval may hinder subsequent analogical thinking. *PloS one*, *14*(2), e0211881. https://doi.org/10.1371/journal.pone.0211881

*van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory*, *25*(9), 1201-1216. https://doi.org/10.1080/09658211.2017.1282518

*van Schie, K., Geraerts, E., & Anderson, M. C. (2013). Emotional and non-emotional memories are suppressible under direct suppression instructions. *Cognition & Emotion*, *27*(6), 1122-1131. https://doi.org/10.1080/02699931.2013.765387

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychology and aging*, *7*(2), 242. https://doi.org/10.1037/0882-7974.7.2.242

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Waldhauser, G. T., Lindgren, M., & Johansson, M. (2012). Intentional suppression can lead to a reduction of memory strength: behavioral and electrophysiological findings. *Frontiers in psychology*, *3*, 401. https://doi.org/10.3389/fpsyg.2012.00401

*Waldhauser, G. T., Johansson, M., Bäckström, M., & Mecklinger, A. (2011). Trait anxiety, working memory capacity, and the effectiveness of memory suppression. *Scandinavian Journal of Psychology*, *52*(1), 21-27. https://doi.org/10.1111/j.1467-9450.2010.00845.x

*Wang, Y., Cao, Z., Zhu, Z., Cai, H., & Wu, Y. (2015). Cue-independent forgetting by intentional suppression–Evidence for inhibition as the mechanism of intentional forgetting. *Cognition*, *143*, 31-35. https://doi.org/10.1016/j.cognition.2015.05.025

Warren, S. L., Heller, W., & Miller, G. A. (2021). The structure of executive dysfunction in depression and anxiety. *Journal of Affective Disorders*, *279*, 208-216. https://doi.org/10.1016/j.jad.2020.09.132

Weller, P. D., Anderson, M. C., Gómez-Ariza, C. J., & Bajo, M. T. (2013). On the status of cue independence as a criterion for memory inhibition: Evidence against the covert blocking hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1232. https://doi.org/10.1037/a0030335

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual review of psychology*, *51*(1), 59-91. https://doi.org/10.1146/annurev.psych.51.1.59

*Wessel, I., Huntjens, R. J., & Verwoerd, J. R. (2010). Cognitive control and suppression of memories of an emotional film. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(2), 83-89. https://doi.org/10.1016/j.jbtep.2009.10.005

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological review*, *120*(2), 329. https://doi.org/10.1037/a0031542

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method–directed forgetting. *Cerebral Cortex*, *18*(3), 670-682. https://doi.org/10.1093/cercor/bhm101

*Yuan, S., Ding, C., Yang, W., Xu, M., Zhang, L., Yao, X., ... & Yang, D. (2018). Gray matter volume of the dorsolateral prefrontal cortex moderates the relationship between rumination and depressed mood. *Current Psychology*, 1-10. https://doi.org/10.1007/s12144-018-0048-8

Zellner, M., & Bäuml, K. H. (2006). Inhibitory deficits in older adults: List-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 290. https://doi.org/10.1037/0278-7393.32.3.290

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. https://doi.org/10.1038/srep37556

*Zhang, R., Yang, X., Yang, R., Xu, Z., Sui, N., & Gao, X. (2020). Wanting to eat matters: Negative affect and emotional eating were associated with impaired memory suppression of food cues. *Appetite*, *150*, 104660. https://doi.org/10.1016/j.appet.2020.104660

*Zhu, Z., & Wang, Y. (2020). Forgetting unrelated episodic memories through suppression-induced amnesia. *J. Exp. Psychol. Gen*. http://dx.doi.org/10.1037/xge0000782