

*Forgetting as a consequence of retrieval:  
a meta-analytic review of retrieval-induced  
forgetting*

Article

Accepted Version

Murayama, K., Miyatsu, T., Buchli, D. and Storm, B. C. (2014)  
Forgetting as a consequence of retrieval: a meta-analytic  
review of retrieval-induced forgetting. *Psychological Bulletin*,  
140 (5). pp. 1383-1409. ISSN 1939-1455 doi:  
<https://doi.org/10.1037/a0037505> Available at  
<http://centaur.reading.ac.uk/37052/>

It is advisable to refer to the publisher's version if you intend to cite from the  
work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1037/a0037505>

Publisher: American Psychological Association

All outputs in CentAUR are protected by Intellectual Property Rights law,  
including copyright law. Copyright and IPR is retained by the creators or other  
copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

**Forgetting as a Consequence of Retrieval:  
A Meta-Analytic Review of Retrieval-Induced Forgetting**

Kou Murayama

University of Reading

Toshiya Miyatsu

Washington University in St. Louis

Dorothy Buchli

University of California, Los Angeles

Benjamin C. Storm

University of California, Santa Cruz

*Accepted by Psychological Bulletin*

**Author Note**

Correspondence should be addressed to Benjamin C. Storm, Department of Psychology, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064. E-mail: bcstorm@ucsc.edu

We thank Drs. Robert A. Bjork, Elizabeth A. Bjork, and other members of CogFog for helpful comments regarding this research.

### **Abstract**

Retrieving a subset of items can cause the forgetting of other items, a phenomenon referred to as *retrieval-induced forgetting*. According to some theorists, retrieval-induced forgetting is the consequence of an inhibitory mechanism that acts to reduce the accessibility of non-target items that interfere with the retrieval of target items. Other theorists argue that inhibition is unnecessary to account for retrieval-induced forgetting, contending instead that the phenomenon can be best explained by non-inhibitory mechanisms, such as strength-based competition or blocking. The current paper provides the first major meta-analysis of retrieval-induced forgetting, conducted with the primary purpose of quantitatively evaluating the multitude of findings that have been used to contrast these two theoretical viewpoints. The results largely supported inhibition accounts, but also provided some challenging evidence, with the nature of the results often varying as a function of how retrieval-induced forgetting was assessed. Implications for further research and theory development are discussed.

*Keywords:* retrieval-induced forgetting, inhibition, interference, meta-analysis, retrieval practice paradigm, part-list cuing

When people conceptualize how information is stored and retrieved, they often assume that memory acts in a way that is analogous to a computer or recording device. Specifically, they assume that information that is deemed worth remembering is encoded or ‘recorded’ for later retrieval, and that when such information is retrieved, it will exist just as it was, unaltered by the retrieval process. Research suggests, however, that the dynamics of memory are decidedly more complex than that. The very act of retrieval can alter the accessibility of information in memory, such that items that are retrieved become more recallable in the future than they would have been otherwise, and related information that was not retrieved becomes less recallable. In this way, retrieval modifies memory with both positive and negative consequences (e.g., Bjork, 1975). The negative consequence, referred to as *retrieval-induced forgetting* (Anderson, Bjork, & Bjork, 1994), is the focus of the present meta-analysis.

Studies of retrieval-induced forgetting have typically employed some variant of a retrieval-practice paradigm consisting of three phases: study, retrieval practice, and final test. During the study phase, participants are presented with a series of category-exemplar pairs drawn from a number of different categories (e.g., *fruits-orange*, *fruits-lemon*, *drinks-rum*, *drinks-vodka*). The pairs are usually presented individually for several seconds in a semi-randomized order and participants are instructed to either study the pairs for a subsequent test or simply think about the association between the categories and their exemplars. During the retrieval-practice phase, participants are guided to retrieve half of the exemplars from half of the categories. The experimenter is able to select which items receive practice by providing category-plus-item-specific cues (e.g., *fruit: or \_\_\_*) that uniquely identify a subset of the exemplars. Often, participants will undergo several rounds of retrieval practice, attempting to retrieve the same exemplars several times each. After a brief delay, typically filled with some sort of distractor task, participants are tested on their ability to recall the exemplars.

The retrieval-practice paradigm creates three types of items: Rp+, Rp-, and Nrp. Rp+ items refer to practiced exemplars (i.e., *orange*); Rp- items refer to non-practiced exemplars from practiced categories (i.e., *lemon*); and Nrp items refer to exemplars from non-practiced categories (i.e., *rum*, *vodka*). Typically, two findings emerge. First, Rp+ items are better recalled than are both Rp- and Nrp items, an observation that replicates work on the positive consequences of retrieval (Landauer & Bjork, 1978; Roediger & Butler, 2011; Roediger & Karpicke, 2006). Second, Rp- items are recalled *less well* than are Nrp items, indicating that remembering some items during retrieval practice causes participants to forget other items from the same practiced categories. This difference in recall performance between non-practiced items from practiced categories (i.e., Rp- items) and non-practiced items from non-practiced categories (i.e., Nrp items) is referred to as retrieval-induced forgetting (RIF).

The phenomenon of RIF has attracted considerable empirical attention, and indeed, several extensive narrative reviews of the literature have been published (e.g., Anderson, 2003; Bäuml, 2007; Raaijmakers & Jakab, 2013; Storm & Levy, 2012; Verde, 2012). Presently, however, no meta-analyses have been conducted on the phenomenon. This is surprising given the large empirical literature. In fact, nearly 200 articles related to retrieval-induced forgetting have been published to date, and interest in the phenomenon continues to grow (Figure 1). We believe that the quantitative integration of the past literature has the potential to sketch a more nuanced picture of the theoretical mechanisms and boundary conditions associated with RIF in ways that would be difficult to achieve through a qualitative survey of the literature. Thus, the

purpose of the present article is to report and summarize key findings from the first meta-analysis on retrieval-induced forgetting.

### **Broad Theoretical Perspectives: Inhibition-based Account and Competition-based Account**

To begin, it is important to emphasize that retrieval-induced forgetting refers to the empirical phenomenon of retrieval practice causing non-practiced items to become less recallable; it does not specify the theoretical mechanism by which non-practiced items become less recallable (Storm & Levy, 2012). Several mechanisms have been argued to account for the phenomenon of RIF, and although a number of distinct theories have been put forth, the majority can be grouped into two broad categories: *inhibition-based forgetting theories* and *competition-based forgetting theories*.

Anderson and colleagues have proposed that the primary mechanism underlying retrieval-induced forgetting is inhibition (Anderson, 2003; Levy & Anderson, 2002; see also Bäuml, 2007; Bjork, Bjork, & Caughey, 2007; C. M. MacLeod & Hulbert, 2011; M. D. MacLeod & Saunders, 2008; Norman, Newman, & Detre, 2007; Storm, 2011; Storm & Levy, 2012). According to this perspective, attempts to retrieve a target item from memory cause multiple items associated with the same retrieval cue to become activated. This activation creates competition, and in order to selectively retrieve the target item, the nontarget items are inhibited. In this way, inhibition functions to reduce interference from nontarget items during retrieval practice. For example, when the retrieval-practice cue, *fruit: or \_\_*, is presented, *lemon* may become inappropriately activated, and to facilitate the retrieval of *orange*, *lemon* is inhibited, thus rendering it less accessible on the subsequent final test.

A variety of perspectives on the inhibition account have been put forth. For example, Anderson (2003) has argued that inhibition acts at the level of the nontarget item's representation, making it less recallable even when tested using independent cues (Anderson & Spellman, 1995; see also Bäuml, 2007). Others, however, have suggested that inhibition may act in a cue-specific way to reduce the accessibility of items specifically in relation to the cues that inappropriately activated them (see e.g., Perfect et al., 2004; Storm & Levy, 2012). Furthermore, some have argued that inhibition is accomplished via frontally-mediated executive-control processes (e.g., Levy & Anderson, 2002; Roman, Soriano, Gomez-Ariza, & Bajo, 2009; see also Marsh, Sörqvist, Beaman, & Jones, 2013), whereas others have argued that inhibition could be accomplished more locally, such as in the medial temporal lobe (see Norman et al., 2007, though the model does predict that executive control could be involved in the selection process). A common theme among all these perspectives, however, is that forgetting is thought to be the consequence of an adaptive or goal-directed process that facilitates retrieval by making interfering information less accessible.

Other researchers have argued that inhibition is unnecessary to explain retrieval-induced forgetting (e.g., Dodd, Castel, & Roberts, 2006; Jonker, Seli, & MacLeod, 2013; C. M. MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003; Perfect et al., 2004; Raaijmakers & Jakab, 2013, Verde, 2012), and that the forgetting effect can be sufficiently explained as the consequence of increased competition or inappropriate retrieval cues at final test. Anderson considered many of these non-inhibitory competition-based accounts of RIF in his seminal papers as well (Anderson & Bjork, 1994; Anderson et al., 1994). According to most such accounts, retrieval of a subset of items strengthens those items and causes them to interfere with other non-strengthened items, thus preventing the non-strengthened items from being successfully retrieved. For instance, after retrieval practice of *orange*, that exemplar might become more strongly associated with the

category *fruit*. Then, at test, when the participant attempts to retrieve another exemplar associated with the same category (i.e., *lemon*), *orange* may block the successful retrieval of that exemplar. According to this perspective, it is the strengthening of target items during retrieval practice, and the effect of that strengthening on the accessibility of non-strengthened items at test, that causes retrieval-induced forgetting to be observed. Additionally, Jonker et al. (2013) have recently provided a new wrinkle to competition-based accounts in their contextual-cuing hypothesis, arguing that retrieval practice may also cause forgetting by changing the way participants search for items in response to the final test cues (but see also Miguez, Mash, Polack, & Miller, 2014).

It should be noted that few researchers in the field explicitly deny the role of competition in RIF. Anderson and his colleagues, for example, who are strong advocates for inhibition-based theories, have acknowledged that competition should contribute to RIF to some extent (see Anderson, 2003; Anderson & Levy, 2007). Moreover, it has been argued that inhibition-based theories and competition-based theories are *not* necessarily mutually exclusive (Storm & Levy, 2012). Researchers have diverged, however, on the issue of whether inhibition plays any role in producing RIF. Researchers who support competition-based theories argue that RIF can be ascribed to *purely* competition-based processes, denying any role for inhibition-based processes. Researchers who support inhibition-based theories, on the other hand, argue that RIF can be at least *partially* explained by inhibition-based processes. From this standpoint, it is possible that both inhibition-based and competition-based mechanisms contribute to some degree to all demonstrations of retrieval-induced forgetting; and, furthermore, it is possible that the relative contribution of these mechanisms will vary depending on how retrieval-induced forgetting is assessed (for discussions, see Anderson & Levy, 2007; Storm & Levy, 2012; Verde, 2012).

Relatedly, as noted above, neither the inhibition-based account, nor the competition-based account, represents a unitary theory --- different researchers have posited different theoretical perspectives on how and why inhibition or competition causes forgetting. Therefore, the rejection of one specific form of inhibition-based or competition-based forgetting does not necessarily negate all forms of inhibition-based or competition-based forgetting. Although we will focus on and refer to inhibition-based theories and competition-based theories in our attempt to encompass the two broad theoretical perspectives, as will be clearer in our meta-analysis, a more nuanced interpretation of these theoretical perspectives will be necessary to more fully delineate the mechanisms underlying RIF. Indeed, we believe that the predominant focus on seeking evidence that entirely confirms one theory while disconfirming the other has acted to stymie progress in the field (cf. McGuire, 2013), a problem that we hope to alleviate in part by conducting and reporting the current meta-analysis.

### **Critical Properties to Test in Assessing Theoretical Accounts of RIF**

As stated above, a critical question in the literature has been whether inhibition contributes, at least in part, to RIF. From the standpoint of the competition-based account, this question can be viewed as whether competition-based mechanisms are sufficient to explain all observations of RIF. In an extensive review, Anderson (2003) identified several key properties of retrieval-induced forgetting that he believed uniquely support a role for inhibition: cue-independence, retrieval specificity, strength independence, and interference dependence. According to Anderson, these properties indicate that inhibition-based mechanisms are necessary, at least in part, to explain RIF. Researchers have devised a number of experimental paradigms to test these properties, which we review briefly below, as well as research examining

populations with presumed inhibitory deficits. It should be noted that our review in this section was designed to simply introduce each property, not to cover them in any sort of comprehensive way (for more comprehensive qualitative reviews, see e.g., Anderson, 2003; Raaijmakers & Jakab, 2013; Storm & Levy, 2012; Verde, 2012).

### **Cue Independence**

According to Anderson's (2003) account of inhibition-based forgetting, RIF is caused by the direct suppression of the representation of the competing memory trace. As such, RIF should not only be observed with tests that employ the original category cue, but also with tests that employ independent retrieval cues (i.e., cues that are not related to the practiced cue or category). If participants study the pairs *fruit:orange* and *fruit:banana*, for example, retrieval practice for *fruit:orange* should cause the forgetting of *banana* regardless of whether *banana* is tested with the original cue (i.e., *fruit*) or with an independent cue (e.g., *monkey*). In contrast, competition-based theories predict that using an independent cue should eliminate RIF because the strengthening caused by retrieval practice is cue specific, thus allowing other cues not associated with that strengthening to remain unaffected. Thus, according to non-inhibitory accounts, the fact that *orange* is strengthened in association to the cue *fruit* should not make *monkey* less likely to cue *banana*.

Anderson and Spellman (1995) developed the independent-probe technique to test the cue independence of RIF. The independent-probe technique has taken two general forms. The first version is an extralist cuing paradigm, which is almost identical to the normal paradigm, except that the final test consists of novel cues that are independent from those used during study or retrieval practice. The second version is a cross-category cuing paradigm, which differs from typical RIF studies in that the studied categories are designed to be related in ways that lead inhibition to act on items associated with non-practiced categories. In Anderson and Spellman's original study, for example, participants studied *blood* and *tomato* under the *red* category and *radish* and *cracker* under the *food* category. *Tomato* was studied in the *red* category, but *tomato* is also a food. *Radish* was studied under the *food* category, but it is also a red thing. Cross-category RIF is observed when retrieval practice on items like *red:blood* results in the forgetting of other red items (i.e. *radish*) that were studied and tested in association with non-practiced category cues (i.e., *food*). Note that this procedure ensures that the critical cue (i.e., *food*) is independent without using a novel (extralist) cue, as the cue was not practiced in the retrieval practice phase.

### **Retrieval Specificity**

A second characteristic of inhibition-based forgetting theories is that they consider RIF to be retrieval specific. That is, the diminished recall of Rp- items is presumed to be caused by inhibition that takes place during retrieval practice. Accordingly, RIF should only occur if retrieval practice trials actually involve active attempts at retrieval. Without retrieval, there is no need for non-target items to be inhibited, and thus no reason to observe forgetting. Competition-based theories, on the other hand, assume that because forgetting is caused by competition at final test, strengthening a subset items in ways other than retrieval practice should also impair the recall of non-strengthened items from the same categories.

In one of the first studies to test the property of retrieval specificity, Ciranni and Shimamura (1999; Experiment 5) replaced retrieval practice with a restudy session in which participants restudied intact cue-target pairs instead of retrieving the targets. As the restudy condition was successful in strengthening the cue-target associations (in fact, performance of

Rp+ items was comparable between the restudy and retrieval practice conditions), the absence of RIF in the restudy condition has been argued to support the retrieval-specificity property.

Later work by Anderson, Bjork, and Bjork (2000) provided additional evidence of retrieval specificity by manipulating the target of retrieval during the retrieval practice. Although normal retrieval-induced forgetting was observed when participants retrieved a subset of the exemplars during retrieval practice (e.g., *fruit:or* \_\_\_\_\_), no forgetting was observed when participants retrieved a subset of the category names (e.g., *fr*\_\_\_\_: *orange*). This category-retrieval paradigm made it possible to strengthen the cue-exemplar associations without having the participants actually retrieve the exemplars, thus presumably precluding the need for the non-practiced exemplars to be inhibited. The fact that no forgetting was observed has been argued to support the retrieval-specificity property of RIF.

### **Strength Independence**

The strength-independence property is closely related to the retrieval-specificity property in that evidence of retrieval specificity is often interpreted as evidence of strength independence. Strength independence refers to the idea that the degree to which practiced items are strengthened fails to predict the degree of retrieval-induced forgetting. The absence of RIF in the restudy or category-retrieval paradigms is considered to support strength independence because such observations indicate that Rp+ items can be strengthened without causing the impairment of Rp- items. Indeed, looking across the literature, Storm and Levy (2012) noted that there are numerous examples of this sort of independence. Specifically, RIF is routinely observed in instances where Rp+ items are not substantially strengthened, and RIF is often not observed in instances where Rp+ items are substantially strengthened. If forgetting is not directly tied to the strengthening of practiced items then such a finding would pose problems for purely competition-based theories.

One way that researchers have attempted to assess strength independence within a given study is to investigate the relationship between RIF and the facilitation effect observed for Rp+ items at final test (i.e., memory performance of the Rp+ items as compared to baseline Nrp items). If RIF is strength dependent then a positive correlation should be observed. If RIF is strength independent, however, then the extent to which Rp+ items are strengthened should fail to predict the extent to which Rp- items are forgotten, and thus a positive correlation should not be observed. A handful of studies have tested for this correlation within individual studies (e.g., Hanslmayr, Staudigl, Aslan, & Bäuml, 2010; Hulbert, Shivde, & Anderson, 2012; Staudigl, Hanslmayr, & Bäuml, 2010); in the present meta-analysis, we will test for it across studies.

Another way in which strength independence has been tested is through the use of an impossible retrieval-practice paradigm (e.g., Storm, Bjork, Bjork, & Nestojko, 2006). In these studies, participants study a list of category-exemplar pairs and then engage in a modified retrieval-practice task with two conditions: impossible retrieval practice and possible retrieval practice. Items in the possible condition consist of category-plus-two-letter-stem cues with the initial letters of exemplars associated with the category (e.g. *fruit:or*). Items in the impossible condition consist of categories with letter stems that do not match any items related to the category (*fruit:wo*). Because impossible retrieval practice is presumed to strengthen the associations between category and exemplars to a lesser extent than possible retrieval practice, it stands to reason, according to the competition-based accounts, that RIF should be significantly reduced. According to inhibition-based accounts, however, because RIF is strength independent, failed retrieval attempts should cause just as much forgetting as successful retrieval attempts.

### **Interference Dependence**

According to inhibition-based theories, the degree to which nontarget items interfere during retrieval practice should determine the degree to which those items suffer forgetting. One way interference dependence has been tested is by manipulating the taxonomic frequency of the exemplars. For example, Anderson et al. (1994, Experiment 3) compared RIF for exemplars of low taxonomic frequency (e.g., *fruit: fig, guava*) with that of the exemplars of high taxonomic frequency (e.g. *fruit: banana, orange*). The authors argued that because items of low taxonomic frequency were weakly associated to the retrieval practice cues, they should have been less likely to cause interference during retrieval practice, and thus, less likely to suffer RIF. This was exactly what Anderson et al. observed. Low-frequency exemplars suffered significantly less forgetting than did high-frequency exemplars.

Interference dependence has been tested in a number of other ways as well (e.g., Bäuml & Sameni, 2010; Jakab & Raaijmakers, 2009; Levy, McVeigh, Marful, & Anderson, 2007; Shivde & Anderson, 2001; Storm & Angello, 2010; Storm, Bjork, & Bjork, 2007), but owing to differences in methodology, and the small number of studies using each methodology, these lines of evidence do not lend themselves well to meta-analysis. One way that some researchers have argued that interference dependence can be assessed, however, that does lend itself well to meta-analysis, is examining the relative amount of RIF observed using ad-hoc or episodically defined categories (e.g., Jonker et al., 2013; Raaijmakers & Jakab, 2013; Verde, 2012). These types of categories have been used routinely to test RIF (e.g., Saunders & MacLeod, 2006; Shaw, Bjork, & Handal, 1995; Storm, Bjork, & Bjork, 2005). Saunders and MacLeod (2006), for example, asked participants to study narratives containing information about two separate burglaries. In their experiment, a house owner's name (e.g., Thompson) served as a category cue and objects in that house served as targets. If ad-hoc categories are relatively weak in terms of their pre-existing associations (like low taxonomic frequency category exemplars), then they might be expected to cause less interference during retrieval practice and thus suffer less RIF (but for a discussion of problems with this assumption, see the General Discussion).

### **Clinical Populations and Retrieval-induced Forgetting**

Several clinical populations have been argued to suffer impairments in the executive control processes of inhibition, such as individuals with frontal lobe damage (Stuss & Benson, 1986), clinical depression (Mayberg, 1994), attention deficit hyperactivity disorder (ADHD; Barkley, 1997), and schizophrenia (Barch, 2009; see also Sterkaj, 2012). In addition, older adults (Hasher & Zacks, 1988) and younger children (Bjorklund & Harnishfeger, 1990) have been argued to have deficits in executive control. These populations allow for a potentially valuable assessment of theoretical accounts of RIF. If RIF does reflect individual differences in general inhibitory functioning, then individuals with established inhibitory deficits, such as those described above, should exhibit reduced forgetting. This prediction has been examined in a number of ways (e.g., Aslan & Bäuml, 2010, 2012; Conway & Fthenaki, 2003; Ford, Keating, & Patel, 2004; Nestor et al., 2005; Soriano, Jimenez, Roman, & Bajo, 2009; Storm & White, 2010; Zellner & Bäuml, 2005).

### **Other Critical Elements in Research on RIF**

Given the different theoretical perspectives explaining RIF, cue independence, retrieval specificity, strength independence, interference dependence, and individual differences have been central topics in recent discussions of RIF (Anderson, 2003; Bäuml, 2007; Raaijmakers & Jakab, 2013; Storm & Levy, 2012; Verde, 2012). There are, however, other critical elements that

deserve attention. These aspects will be reviewed and discussed below.

### **Controlling for Output Interference at Final Test**

We believe that whether output order is controlled at final test in a given study is critical for interpreting the theoretical mechanism(s) underlying a given effect of RIF. When output order is not controlled, participants may recall strengthened Rp+ items first, thus causing the forgetting of non-strengthened Rp- items via *output interference* (Roediger, 1974; A. D. Smith, 1973; Tulving & Arbuckle, 1966). It is important to emphasize that output interference and RIF are intrinsically related to each other. Indeed, output interference can be considered a form of RIF that occurs at final test in that it reflects a decline in the accessibility of some items as the result of the retrieval of other items. Importantly, as with the term retrieval-induced forgetting, output interference refers to an empirical phenomenon that does not specify the underlying mechanism, and may therefore be explained by both competition-based and inhibition-based mechanisms (see Bäuml, 1998). Regardless of why output interference is observed, failing to control for it creates a problem by making it difficult, if not impossible, to determine the mechanism underlying RIF effects observed using the retrieval-practice paradigm. For example, if Rp- items suffer more output interference at test than Nrp items, then RIF may be observed at final test even if Rp- items were not actually inhibited during retrieval practice. Furthermore, if output interference has a significant competition-based component—and there are reasons to think that it might—then the problems associated with not controlling for output interference can become magnified, as participants may exhibit greater RIF effects that are a consequence of competition at test, not inhibition during retrieval practice.

Output interference is likely to occur when the final memory test is structured in such a way that allows participants to determine the order in which items are recalled. For example, when category-cued recall tests are employed, participants are given each category name and asked to recall all of the associated exemplars that were studied. On such tests, participants may recall Rp+ items first, thus causing Rp- items to suffer greater levels of output interference than Nrp items. Fortunately, output interference can be controlled by specifying the order of output, forcing participants to selectively target Rp- and Nrp items in a pre-determined order that is not confounded by the testing of Rp+ items. For example, the use of item-specific category-plus-stem cues at test allows researchers to control the order in which items are recalled. Some researchers have participants recall Rp- items and a matched set of Nrp items first, whereas others have participants recall all items in a randomized order. Both practices prevent participants from selectively recalling Rp+ items before Rp- items.

Because output interference may confound competition-based and inhibition-based forgetting, the failure to control for it may make it very difficult, if not inappropriate, to test certain theoretical predictions. Thus, we examined many factors and moderating variables as a function of whether output interference was controlled, with the assumption that studies controlling for output interference would provide a clearer assessment of the theoretical processes underlying RIF.

### **Materials**

Retrieval-induced forgetting is most commonly examined with category-exemplar word pairs. According to both inhibition-based theories and competition-based theories, however, RIF should be a more general effect that can be generalized to a wide range of memory-related phenomena. In fact, RIF or RIF-like phenomena have been observed using text passages (Little, Storm, & Bjork, 2011), lexical categories (Bajo, Gomez-Ariza, Fernandez, & Marful, 2006),

autobiographical memories (Barnier, Hung, & Conway, 2004), pictures (Ford et al., 2004), videos (Miguelles & Garcia-Bajos, 2007), factual propositions (Anderson & Bell, 2001), social conversations (Coman, Manier, & Hirst, 2009), insight problems (Storm, Angello, & Bjork, 2011), divergent thinking tasks (Storm & Patel, 2014), personality traits (Dunn & Spellman, 2003), motor actions (Tempel & Frings, 2013), visual objects (Ciranni & Shimamura, 1999), arithmetic facts (Campbell & Thompson, 2012), semantic memory (Johnson & Anderson, 2004), goals (McCulloch, Aarts, Fujita, & Bargh, 2008), scripts (Garcia-Bajos & Miguelles, 2013), memory for self-performed actions (Sharman, 2011), and spatial locations (Gómez-Ariza, Fernandez, & Bajo, 2012).

### **Final Memory Tests**

The most common method for assessing RIF has been to administer a final cued-recall test (e.g., category-cued or category-plus-stem-cued). Another way researchers have assessed RIF is through the use of an item-recognition test. Unlike cued-recall tests, item-recognition tests involve showing participants the items they studied along with other items from the same categories that were not studied (i.e., lures) and asking participants to determine whether each item had been studied. Although theoretical perspectives on the use of item-recognition tests have been mixed, some researchers have argued that they can provide a way of assessing inhibitory-based forgetting (Gomez-Ariza, Lechuga, Pelegrina, & Bajo, 2005; Hicks & Starns, 2004). Another way RIF has been assessed is through the use of implicit memory tests, which can be defined as tests that do not require participants to deliberately or consciously recollect the previously studied items (Schacter, 1987). Implicit tests can take a variety of forms, including perceptual identification, lexical decision, word-fragment completion, and recognition reaction time (e.g., Bajo, Gómez-Ariza, Fernandez, & Marful, 2006; Butler, Williams, Zacks, & Maki, 2001; Parker & Dagnall, 2009; Perfect et al., 2004).

### **Extralist Paradigm**

A number of studies have examined RIF as a consequence of extralist retrieval practice. In this variant of the paradigm (e.g., Bäuml, 2002; Storm et al., 2006), participants study category-exemplar pairs, but instead of retrieving a subset of those same exemplars during retrieval practice, they are asked to generate new, unstudied items belonging to the same categories, often using the same type of category-plus-stem retrieval cues used in the more typical episodic form of retrieval practice.

One advantage of the extralist paradigm is that it can alleviate, at least in part, the context-cuing effects outlined by Jonker et al. (2013). According to the context-based account, RIF occurs because practiced category cues ( $R_p+$  and  $R_p-$  items) at test activate the retrieval-practice phase context, disrupting the search for items that were not practiced ( $R_p-$  items). Non-practiced category cues (i.e.,  $N_{rp}$  items), on the other hand, activate the study phase context, and accordingly,  $N_{rp}$  items can be recalled without context-based interference. In this way, the context-based account can explain RIF without requiring memory inhibition. Importantly, in most versions of the extralist paradigm, it is obvious to participants (and often they are explicitly told of the fact) that all items to-be-recalled at final test were presented in the study phase and not retrieved during retrieval practice. This awareness should lead participants to attempt to recall items directly from the original study phase context as opposed to the retrieval practice context, thus presumably reducing the extent to which context-cuing causes RIF.

### **Between- and Within-Subjects Design**

Retrieval-induced forgetting has been most commonly tested using a within-subjects

design. That is, Rp+, Rp-, and Nrp conditions are manipulated within participants. It is also possible, however, to include another group of participants who do not engage in any form retrieval practice and to use the memory performance of these participants as a baseline (i.e., a between-subjects Nrp condition; e.g., Shaw et al., 1995). One benefit of comparing RIF studies using within-subjects design with those using between-subjects design is that we can test the potential deflation of baseline memory performance (i.e., memory performance for Nrp items). For example, Tsukimoto and Kawaguchi (2006) argued that items from unpracticed categories (Nrp items) with a within-subjects design may be suppressed during the retrieval-practice phase because practiced and unpracticed categories (i.e., all of the studied categories) can be represented in a common episodic context (see also Anderson et al., 1994). Thus, if baseline deflation does occur, then it may decrease the amount of RIF that is observed in within-subjects designs (see also Kato, 2007).

### **Effects of Delay**

Although most studies of RIF are completed in the context of a single-session one-hour experiment, some studies have examined the consequences of retrieval practice on final test performance using much longer delays, sometimes after 24 hours or even a week (e.g., Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009, 2010; Garcia-Bajos, Migueles, & Anderson, 2009; Saunders, Fernandes, & Kosnes, 2009; Storm, Bjork, & Bjork, 2012; Storm et al., 2006). Theoretical implications for the effects of delay on RIF are debatable. Although some researchers have argued that the inhibition-based account predicts that RIF reflects a temporary or transient reduction in the accessibility of items in memory, expecting it to be diminished or even eliminated after a delay (M. D. MacLeod & Hulbert, 2011; M. D. MacLeod & Macrae, 2001; Raaijmakers & Jakab, 2013), others have argued that under certain conditions it might be possible for inhibition to have persisting consequences (Storm et al., 2012; see also Anderson, 2003, p. 423 for further discussion). Regardless, the temporal boundaries of RIF remain largely unknown, thus limiting our understanding of how retrieval practice affects the recall of other information in the long term.

### **The Current Research**

The purpose of the present research is to conduct a comprehensive meta-analysis of the existing empirical studies on RIF. Importantly, by focusing on the factors discussed above, especially on the factors that are critically relevant to theoretical perspectives (i.e., cue independence, retrieval specificity, strengthen independence, interference dependence, individual differences, and output interference), our meta-analysis aims to further inform the ongoing theoretical debate on the nature of RIF.

### **Method**

#### **Sample of Studies**

The studies included in our meta-analysis were identified via a thorough search of the literature for peer-reviewed studies and dissertations written in English and published before August, 2012. Specifically, we searched PsycINFO and Web of Science using the keyword *retrieval induced forgetting*. To ensure that our meta-analysis was as comprehensive as possible, we also checked the reference sections of major review papers on RIF and searched for published papers by major contributors in the field. In total, 194 papers were identified for possible inclusion in the meta-analysis.

We then screened the pool of studies using the following inclusion criteria. First, the study had to report original empirical data on RIF. Second, the study had to use some variant of the retrieval-practice paradigm involving retrieval practice and a final test<sup>1</sup>. Studies using similar but distinct paradigms (e.g., output interference, part-list cuing, think/no-think) were excluded. All studies that assessed RIF were included, regardless of the specific materials (e.g., pictures, free conversation) or measures (e.g., memory performance, response latency) that were employed. Third, the study had to have an appropriate baseline condition (Nrp). Some studies did not include a pure no-retrieval-practice condition (e.g., Wimber, Rutschmann, Greenlee, & Bäuml, 2009), thus making it difficult to measure RIF in a way that is comparable to other studies. These studies were excluded from the meta-analysis. Fourth, the study had to contain sufficient information to estimate the effect size and standard error of the difference between performance in the Rp- condition and the Nrp condition (see the Calculating and Integrating Effect Sizes section).

In total, our search yielded 143 research papers involving  $k = 512$  separate samples with 759 effect size estimates (some studies utilized a within-subjects design or assessed multiple dependent variables, producing multiple effect size estimates for a single sample). Papers included in our meta-analysis are marked with an asterisk in the reference section.

### Sample Coding

The following information was coded for each effect size estimate. To ensure the reliability and consistency of the coding, we took a two-step procedure. In the first step, the first, second, third, and fourth authors coded the studies. These authors had regular meetings to discuss and elaborate the coding scheme. In the second step, the first author went over all the papers again to make sure they were coded appropriately and consistently.

Our coding focused specifically on moderators likely to be of theoretical importance. As described in the introduction, support for the inhibition account of RIF has come from evidence of several observations, including: cue independence, retrieval specificity, interference independence, strength independence, and individual differences. We coded several moderators relevant to each of these criteria.

**Controlling for Output Interference.** One of the most important moderators in the current meta-analysis is whether or not a study controlled for the order in which items in a category were recalled. As described in the introduction, when order is not controlled, RIF can result from output interference at final test, making it difficult to ascertain whether inhibition had occurred at the time of retrieval practice. Fortunately, the effects of output interference can be reduced or eliminated by employing item-specific cues (e.g., letter-stems, word-fragments, item-recognition, etc.) to control the order in which items are tested, or by presenting cues that selectively target Rp- and Nrp items (and not Rp+ items). Some studies presented item-specific cues in random order, and other studies forced participants to recall Rp- items before Rp+ items. In either case, we coded output interference as being controlled. It should be noted that some samples forced participants to recall Rp+ items before Rp- items to directly examine the effects of output interference (e.g., Anderson et al., 1994, Experiment 2). In such cases, which were relatively rare, the samples were coded as not controlling for output interference. Studies that did not test Rp+ items (e.g., Levy et al., 2007), or that did not have Rp+ items (e.g., Storm et al.,

2006), were considered as controlling for output interference because there was no opportunity for Rp+ items to be retrieved before Rp- items at test. Also, studies that examined RIF with independent probes (e.g., Anderson & Spellman, 1995) were coded as controlling for output interference regardless of the final test type used, because in these studies the category cues used to probe Rp- items were always different from those used to probe Rp+ items (i.e., the output of Rp+ items should not interfere with the output of Rp- items). In the end, we were left with a sample of studies that controlled for output interference versus a sample that did not.

**Cue-independence.** We coded whether independent cues/probes were employed during retrieval practice or final test (Anderson & Spellman, 1995). Additional coding distinguished between cue-independent paradigms that used a cross-category paradigm and those that used an extra-list cuing paradigm. There were some variants of the cross-category paradigm (e.g., first-order vs. second-order cross-category RIF; see M. D. MacLeod & Saunders, 2005), but given the small number of these studies, we coded them as a single sub-category.

**Retrieval specificity and strength independence.** Two experimental paradigms were coded to examine the retrieval-specificity and strength-independence assumptions of the inhibitory account of RIF: (1) The restudy paradigm, in which participants simply read or restudied a subset of the studied material in place of what would normally be retrieval practice (e.g., Ciranni & Shimamura, 1999); and (2) the category-retrieval paradigm, in which participants retrieved the category name given its associated exemplar (e.g., Anderson et al., 2000). Each of these experimental approaches has been used to test the assumption that retrieval of Rp+ exemplars during retrieval practice is necessary to cause forgetting (retrieval specificity), as well as the assumption that RIF is not directly tied to the extent to which practiced items are strengthened (strength independence).

Strength independence was also examined by quantifying the extent to which Rp+ items were strengthened by retrieval practice. First, we coded performance levels during retrieval practice. Second, we coded the effect size of the facilitation effect observed at final test resulting from retrieval practice (the standardized difference between Rp+ and Nrp memory performance). The effect size was computed in the same way that RIF was computed.

We also coded the presence or absence of feedback during retrieval practice, as according to the inhibition account feedback during retrieval practice should increase the strengthening of Rp+ without leading to larger effects of retrieval-induced forgetting. Competition-based accounts, however, predict that additional strengthening of Rp+ items through feedback should lead to larger effects of retrieval-induced forgetting. Finally, we coded studies using the impossible retrieval paradigm in which impossible retrieval cues were provided, thus ensuring that participants failed to retrieve any items during retrieval practice (e.g., Storm et al., 2006).

**Interference dependence.** To evaluate interference dependence, we focused on the samples that directly compared the RIF effect for strong versus weak exemplars (in terms of taxonomic frequency; e.g., Anderson et al., 1994). We did not code studies that did not directly manipulate category-exemplar association strength in order to ensure the comparability of these two conditions in our meta-analysis. We also sought to evaluate interference dependence by coding whether studies used to-be-learned materials with no pre-existing associations between categories and exemplars (i.e., episodically defined categories; e.g., Shaw et al., 1995).

**Individual differences.** We coded the samples that examined older adults, young children, and clinical patients assumed to have deficits in executive functioning (i.e., frontal lobe lesion patients, clinical depression patients, ADHD, schizophrenic patients, and posttraumatic stress disorder patients; inclusion criteria was determined prior to the analysis). The clinical patients were coded together given that there were not a sufficient number of samples for each specific clinical subgroup. Most of these studies included a control condition (young adults as a control for older adults or young children; healthy adults as a control for clinical patients), and for the purpose of comparison, these control samples were also coded.

**Other moderators of interest.** Several additional moderator variables were coded. First, we coded the type of study materials, which included lists of words, personal traits, visual stimuli, actions, autobiographical memories, factual propositions, and text passages. Second, we coded the final test type, which included category-cued recall, category-plus-stem cued recall, item recognition, and implicit memory tasks such as perceptual identification and recognition latency. Third, we coded studies using an extralist paradigm in which retrieval practice consisted of generating items that were not presented during the study phase (Bäuml, 2002; Storm et al., 2006). Fourth, we coded whether Nrp and Rp- conditions were compared as a within-subjects or a between-subjects factor. Fifth, we coded the length of the delay between retrieval practice and final test to address the potentially transient nature of RIF (M. D. MacLeod & Macrae, 2001), focusing specifically on whether or not the test took place on the same day of the experiment (Baran, Wilson, & Spencer, 2010). Finally, we coded the amount of time each item was studied during the study phase and the amount of time participants had to retrieve items during the retrieval-practice phase.

### Calculating and Integrating Effect Sizes

All data analyses were conducted by a combinational use of *R* (we utilized the package *meta* and *metaphor*; Schwarzer, 2012; Viechtbauer, 2010) and Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2008; for technical details, see Borenstein, Hedges, Higgins, & Rothstein, 2009).

**Effect size computation.** The focal effect size in this meta-analysis is a standardized mean difference between Rp- and Nrp conditions. According to Cohen (1988), the standardized mean difference of 0.2 to 0.3 reflects a small effect size, 0.5 reflects a medium effect size, and 0.8 or larger reflects a large effect size. We computed effect sizes using Hedges and Olkin (1985)'s unbiased estimator  $g_j$ , given by

$$g_j = \left( \frac{\bar{Y}_j^{Nrp} - \bar{Y}_j^{Rp-}}{S_j} \right) \times c_j(df)$$

where  $\bar{Y}_j^{Nrp}$  and  $\bar{Y}_j^{Rp-}$  are the sample means for the Nrp and Rp- conditions in the  $j$ th sample.  $S_j$  is the pooled standard deviation of the Nrp and Rp- conditions (Cohen, 1988). As can be seen, the effect sizes were calculated such that a positive effect indicated the existence of an impairing effect of retrieval practice (i.e., RIF effect). When we computed the effect sizes for response

latency,  $\bar{Y}_j^{Nrp} - \bar{Y}_j^{Rp-}$  was replaced with  $\bar{Y}_j^{Rp-} - \bar{Y}_j^{Nrp}$  so that a positive effect also indicated RIF.  $c_j(df)$  is a bias correction factor in the  $j$ th sample and computed by the following formula.

$$c_j(df) = 1 - \frac{3}{4df - 1}$$

When the data were from a between-subjects design with the total sample size =  $N$ , then  $df = N - 2$ . When the data were from a within-subjects design with the sample size =  $N$  (i.e.,  $N$  for each cell),  $df = N - 1$ .

One complication of the current meta-analysis is that RIF has been investigated using both within-subjects and between-subjects designs. This poses several issues (Morris & DeShon, 2002). First, there are different formulae for computing the pooled standard deviation ( $S_j$ ) in a within-subjects design. Several researchers have argued for the use of the standard deviation of the difference score (e.g., Gibbons, Hedeker, & Davis, 1993), and indeed many of the studies included in our meta-analysis reported effect size estimates based on this definition. However, as discussed by Morris and DeShon (2002), this effect size metric is not comparable with the effect size estimates obtained from a between-subject design (see also Dunlap, Cortina, Vaslow, & Burke, 1996). Accordingly, we defined the pooled standard deviation using the standard deviation of each group in both between- and within-subject designs<sup>2</sup>. That is, for both designs,  $S_j$  was defined as

$$S_j = \sqrt{\frac{(S_j^{Nrp})^2 + (S_j^{Rp-})^2}{2}}$$

where  $S_j^{Nrp}$  and  $S_j^{Rp-}$  are the standard deviations of the Nrp and Rp- conditions in  $j$ th sample, respectively. In the few between-subject cases where the sample size for each condition was different, we estimated  $S_j$  using a weighted average (by sample size) of the standard deviations following the formula provided by Cohen (1988).

A second complication for including both within-subject and between-subject designs is that the statistics necessary to estimate the effect size is different. If means and standard deviations for Nrp and Rp- conditions are available, effect size can be computed for both designs (assuming that sample size information is also available). On the other hand, when only  $t$  or  $F$  values (or associated  $p$  values) comparing Nrp and Rp- conditions are available, effect sizes can be estimated in a between-subject design (Lipsey & Wilson, 2001), but not in a within-subject design. To compute effect size from these test statistics in a within-subject design, the correlation between Nrp condition and Rp- condition is required (Dunlap et al., 1996). Unfortunately, none of the studies included in our meta-analysis reported this correlation. However, it is possible to estimate the correlation if a study with a within-subject design includes means, standard deviations, and  $t$  or  $F$  test statistics (or associated  $p$  values). Accordingly, we first conducted a meta-analysis of the correlation between Nrp and Rp- conditions (see Hunter & Schmidt, 1990) using the studies that reported all relevant information ( $k = 64$ ), and then used the estimated correlation ( $r = 0.43$ , 95% CI [0.37, 0.50]) to compute the within-subject effect size estimates of those studies that reported only test statistics.

Third, sampling variance estimates ( $v_j$ ) are also influenced by study design. Morris and DeShon (2002) provided separate formulae for computing sampling variance for within-subject and between-subject designs, and we followed these formulae (see also Noortgate & Onghena, 2003). To compute the sampling variability of studies with within-subjects designs, the correlation between Nrp and Rp- condition was again needed. We used the same estimate obtained above to compute the sampling variability of these studies.

Several studies only reported the presence or absence of statistical significance (e.g., “the results were not statistically significant”) or the range of a particular statistical value (e.g., “ $F < 1$ ”). Given that most of these cases represent RIF effects that did not reach statistical significance, excluding these studies could potentially overestimate the integrated RIF effect sizes. To avoid such artificial inflation of effect size estimates, for these studies, we randomly generated effect size estimates within the range of the reported information. For example, if a sample reported the RIF effect as  $t$  or  $F < 1.00$ , an effect size estimate was calculated by making  $F$  equal to a number between 0.00 and 1.00 produced by a random number generator (this is akin to stochastic imputation procedure to handle missing data without biasing point estimates, Enders, 2010; see Cameron & Pierce, 1994, for a similar procedure). When the direction of the effect was available from a table or figure, this information was also considered. If direction information was not available then the direction of the effect was determined randomly as well. There were also studies that investigated the difference in RIF between conditions (e.g., is RIF different in condition A vs. condition B?). In some of these studies, researchers did not find a significant interaction between the RIF factor (Nrp vs. Rp-) and the conditions of interest, and they only reported statistics involving the main effect of the RIF factor, but did not report the RIF effect for each condition. For these studies, we provided the same effect size estimates (computed by the reported statistics) for these conditions.

**Integration of the effect sizes.** We adopted a random-effects framework to integrate the computed effect sizes (Hedges & Vevea, 1998). In this approach, the overall point estimate of the effect sizes is obtained by computing the weighted average of the effect sizes ( $g_j$ ) with the weight of  $j$ th sample  $w_j$  given by

$$w_j = \frac{1}{v_j + \tau^2}$$

where  $\tau^2$  is the between-sample variance (the variance of the effect size parameters across the population of samples), estimated by the method of moments method. The standard error of the averaged effect sizes  $V$  is estimated as follows:

$$V = \frac{1}{\sum_{i=1}^K w_i}$$

where  $K$  is the total number of samples. This value is used to construct a 95% confidence interval (CI) of the average effect size.

As indicated earlier, there were a number of samples that included more than one effect size, either because the studies employed multiple dependent variables (e.g., memory

performance and reaction time) or because they involved multiple within-subject conditions. When an analysis was concerned with an overall effect size, the average effect size within the sample (see Borenstein et al., 2009) was computed and used in the meta-analysis. When an analysis was interested in a specific outcome variable or condition (e.g., a meta-analysis focusing on reaction time), then only the focal effect sizes were selected and included in the meta-analysis.

**Moderator analysis.** The presence of between-sample variance (effect size heterogeneity) was assessed by  $Q$  statistics.  $Q$  statistics test the null hypothesis that all individual effect sizes estimate the same population effect size (Lipsey & Wilson, 2001). Statistically significant  $Q$  statistics indicate the presence of possible moderator variables. Moderator analysis was performed by simply repeating the meta-analysis for the subsamples built by the categories of a moderator variable. When we were interested in the direct comparison of the effect size between two subsamples, we ran a mixed-effects model to test the statistical significance of the effect size difference (Borenstein et al., 2009).

Three notes should be made. First, because  $Q$  statistics do not represent the magnitude of between-sample heterogeneity, they were supplemented by  $I^2$  statistics that represent the ratio (0% - 100%) of true heterogeneity to total variance across the observed effect estimates (Higgins, Thompson, Deeks, & Altman, 2003). Second, mixed-effects models can compare the average effect size of subsamples only when the subsamples are not overlapping. Accordingly, when conducting a mixed-effect model, we excluded samples with more than one dependent variables belonging to different categories of a moderator variable of interest. Accordingly, samples included in a given moderator analysis (i.e., separate meta-analysis based on moderator categories) and those used in a given mixed-effects model analysis had the potential to be slightly different.

Finally, when a moderator was a continuous variable, we conducted a meta-regression analysis (Borenstein et al., 2009) to quantify the relationship between the magnitude of the moderator and the RIF effect. Because meta-regression analysis can include only one effect size estimate per sample, when a sample had more than one effect size estimate, we randomly sampled one of these estimates and included it in the analysis.

## Results

### Global Analysis

We first determined the average size and significance of the RIF effect with all studies included (Table 1). The weighted mean of the effect sizes from the full 512 samples was  $g = 0.35$  with a 95% confidence interval (CI) = [0.32, 0.38]. We computed fail-safe  $N$  (Rosenthal, 1979), which is the number of file-drawer (i.e., nonsignificant) studies that would make the effect size statistically nonsignificant. The results showed that fail-safe  $N$  is 56,111, indicating the robustness of RIF effect against publication bias.

It should be noted that this global meta-analysis includes all samples, even samples in which the RIF effect is supposed to be weak (e.g., low-frequency exemplars), non-existent (restudy instead of retrieval practice), or negative (e.g., semantically-integrated text materials tested after a long delay) from a given theoretical perspective, and thus probably underestimates

the actual size of the average RIF effect. In fact, the analysis also indicated that there is substantial and significant variability in the effect size across samples,  $Q(511) = 1381.2, p < .01, I^2 = 63.0\%$ . Nevertheless, the global analysis indicated that, on average, RIF papers produced a small to moderate effect size, and one that was highly robust (i.e., lower bound of the 95% CI is above Cohen's estimate for small effect sizes). We also conducted a meta-analysis after excluding studies that did not actually involve retrieval practice (i.e., retrieval practice was replaced by some other task, such as restudying, category generation, generating mental imagery, etc.), thus providing an arguably more appropriate measure of "retrieval" induced forgetting ( $k = 472$ ; Table 1). The average effect size remained almost the same,  $g = 0.35, 95\% \text{ CI} = [0.32, 0.38]$ .

### **RIF and the Control of Output Interference at Final Test**

In order to examine the effect of controlling for output interference, the entire sample was divided into two subsamples that either controlled for the potentially interfering recall of Rp+ items at final test or did not (Table 1). Note that the total number of samples does not match the sum of the samples in the two output interference conditions because some samples had more than one dependent variable that fall into different output interference categories (this point applies to all of the following analyses on output interference as well). The results clearly indicated that failing to control output interference increased the overall RIF effect: studies not controlling output interference ( $k = 256$ ) showed a larger RIF effect,  $g = 0.50, 95\% \text{ CI} = [0.45, 0.55]$  than those that did ( $k = 280$ ),  $g = 0.22, 95\% \text{ CI} = [0.19, 0.26]$ . This difference was statistically significant,  $Q(1) = 77.0, p < .01$ . The same pattern was observed when limiting our analysis to the studies that involved retrieval practice: studies not controlling output interference ( $k = 238$ ) showed a larger RIF effect,  $g = 0.50, 95\% \text{ CI} = [0.45, 0.54]$  than those that did ( $k = 257$ ),  $g = 0.23, 95\% \text{ CI} = [0.19, 0.27]$ . This difference was statistically significant,  $Q(1) = 71.7, p < .01$ . Importantly, studies that controlled output interference still showed a small, but highly robust, RIF effect (fail-safe  $N = 6008$ ), suggesting that the failure to control for output interference cannot fully explain the total RIF effects observed in the literature.

To make more sense of the effect size estimate obtained above, we conducted a supplementary meta-analysis on the raw mean differences in memory performance (Bond, Wiitala, & Richard, 2003). In this meta-analysis, we only included samples that reported raw memory performance (in percentages; or in a format that enabled us to convert performance into percentages) and associated standard deviations in both Nrp and Rp- conditions. Samples that reported only test statistics were excluded because it is not possible to compute the raw mean difference and sampling variation solely from test statistics. With this set of samples ( $k = 221$ ), the overall standardized effect size ( $g$ ) slightly increased,  $g = 0.40, 95\% \text{ CI} = [0.35, 0.45]$ . Again, studies that did not control output interference ( $k = 142$ ) showed a larger average effect size,  $g = 0.52, 95\% \text{ CI} = [0.45, 0.59]$  than those that did ( $k = 90$ ),  $g = 0.22, 95\% \text{ CI} = [0.16, 0.28]$ . Importantly, the meta-analysis on the raw mean difference showed that the average raw mean difference was 8.5% for the entire sample ( $95\% \text{ CI} = [7.4\%, 9.6\%]$ ), 10.9% for samples that did not control output interference ( $k = 142; 95\% \text{ CI} = [9.4\%, 12.4\%]$ ), and 4.5% for samples that did control output interference ( $k = 90; 95\% \text{ CI} = [3.3\%, 5.6\%]$ ).

Again, we repeated the meta-analysis on raw mean differences after excluding studies that did not actually involve retrieval practice. The results of this more focused analysis were

very similar to those reported above. Specifically, the average raw mean difference was 8.7% for the entire sample ( $k = 193$ , 95% CI = [7.5%, 9.8%]), 10.9% for samples that did not control output interference ( $k = 124$ ; 95% CI = [9.4%, 12.4%]), and 5.0% for samples that did control output interference ( $k = 79$ ; 95% CI = [3.7%, 6.3%]). Taken together, these results indicate that the average RIF effect (i.e., the raw difference in recall performance between Rp- items and Nrp items) is about 9%; 11% when output interference is not controlled and 5% when output interference is controlled. Once again, however, it deserves emphasis that these average effect sizes include studies and manipulations specifically designed to prevent and even reverse the RIF effect. Therefore, it is likely that the effect sizes underestimate the size of the true effects, and we should thus interpret these values with caution (Borenstein et al., 2009).

### **RIF Effect with Different Study Materials**

Table 1 presents the RIF effect across different types of materials. Across all study materials, the overall RIF effect ranged from small to large and all were statistically significant, indicating that RIF is a highly general phenomenon. Moreover, regardless of the type of materials employed, studies that controlled output interference consistently exhibited smaller effect sizes than studies that did not, though the RIF effect was still mostly significant. These findings suggest that regardless of the materials employed, part but not all of a given RIF effect can be explained by output dynamics at test.

One remarkable exception to the above analysis is the very small effect size observed with text materials when output interference is controlled ( $g = 0.08$ , 95% CI = [-0.04, 0.21]). The small effect size may have been produced due to the fact that both retrieval-induced forgetting and retrieval-induced facilitation have been investigated in the majority of experiments using text materials (Chan, 2009, 2010; Chan, McDermott, & Roediger, 2006). Chan (2009), for example, demonstrated that retrieval practice can enhance the later recall of non-practiced text materials when the text is organized in an integrated fashion, particularly after long delays, whereas retrieval practice inhibits the later recall of non-practiced text materials when the integration of text is disrupted by changing the order of sentences. These opposing effects (if they exist) could produce the overall small effect. In fact, relatively high heterogeneity of the effect size was observed in the experiments using text materials ( $I^2$  is between 71.5% - 78.2%).

Accordingly, we computed the average effect size for studies that directly manipulated the integration of the text (and controlled output interference), and compared the effect size of the high-integration condition with that of the low-integration condition. As reported at the bottom of Table 1, the low-integration condition ( $k = 7$ ) showed a significant RIF effect,  $g = 0.25$ , 95% CI = [0.09, 0.42], whereas the high-integration condition ( $k = 7$ ) showed a non-significant RIF effect,  $g = 0.01$ , 95% CI = [-0.15, 0.17]. A mixed-effects model indicated that this difference was statistically significant,  $Q(1) = 4.71$ ,  $p < .05$ . Moreover, within each experimental condition (i.e., low-integration and high-integration), the heterogeneity of the samples was dramatically reduced and not statistically significant,  $Q(6) = 5.95$ , ns,  $I^2 = 0\%$  for the low-integration condition,  $Q(6) = 4.61$ , ns,  $I^2 = 0\%$  for the high-integration condition. These results are consistent with the hypothesis that integration is a key factor moderating RIF with text materials<sup>3</sup>.

### **Testing Theoretical Accounts of RIF**

One of the primary objectives of the current meta-analysis was to examine the effects of moderator variables relevant to theoretical accounts of RIF. In the following sections, the moderation analysis was organized based on several important theoretical predictions.

**Cue-independence.** The effect size of samples utilizing some type of cue-independent paradigm (e.g., cross-category paradigm, extralist cuing paradigm;  $k = 67$ ) was small, but statistically significant,  $g = 0.16$ , 95% CI = [0.09, 0.24] (see Table 2). As confirmed by a mixed-effects model, the RIF effect observed in studies using cross-category paradigms ( $k = 36$ ,  $g = 0.20$ , 95% CI = [0.11, 0.29]) was no different than the RIF effect observed in studies using extralist cuing paradigms ( $k = 37$ ,  $g = 0.18$ , 95% CI = [0.06, 0.30]),  $Q(1) = 0.48$ , ns. However, a separate mixed-effects model revealed that samples that used either of the cue-independent paradigms exhibited a significantly smaller RIF effect than those that did not,  $Q(1) = 18.1$ ,  $p < .01$ .

As discussed earlier, experiments examining the cue-independence of RIF should be immune to output interference, regardless of the test type used, because Rp- items are probed independently of Rp+ items. Consistent with this expectation, the form of the final test failed to influence the magnitude of the cue-independent RIF effect. That is, a mixed-effects model showed that the effect size of samples that employed category-cued independent recall tests ( $k = 24$ ),  $g = 0.18$ , 95% CI = [0.08, 0.28], was not significantly different than that of samples that employed item-specific independent recall tests ( $k = 43$ ),  $g = 0.16$ , 95% CI = [0.06, 0.26],  $Q(1) = 0.08$ , ns.

**Retrieval specificity and strength independence.** As indicated in Table 3, the meta-analysis of two experimental manipulations aimed to address the retrieval-specificity and strength-independence assumptions of RIF provided substantially different results. In the samples using a restudy paradigm ( $k = 17$ ), where participants restudied a subset of items instead of performing retrieval practice for those items, the meta-analysis indicated that restudy did not produce a significant decrease in memory of the items that were not restudied but were from the restudied categories,  $g = 0.07$ , 95% CI = [-0.03, 0.17]. The results indicated non-significant heterogeneity across samples,  $Q(16) = 8.9$ , ns,  $I^2 = 0.0\%$ .

Results from the category-retrieval paradigm ( $k = 7$ ), however, in which participants retrieved the category names when given the category exemplars, exhibited a very different pattern of results. Specifically, the RIF effect in such studies was found to have a small to medium effect size, and the effect was statistically significant,  $g = 0.37$ , 95% CI = [0.04, 0.71]. A closer look at the data indicated a large split between the studies that observed the RIF effect (Jonker & MacLeod, 2012; Raaijmakers & Jakab, 2012) and those that did not (Anderson et al., 1994; Hanslmayr et al., 2010; Saunders et al., 2009). Indeed, there was highly significant heterogeneity across the samples with this paradigm despite the small number of samples,  $Q(6) = 23.0$ ,  $p < .01$ ,  $I^2 = 73.9\%$ . As the effect size was not significantly different in studies controlling output interference versus those that did not, it does not seem that output interference alone can explain this heterogeneity,  $Q(1) = 0.03$ , ns. The results of this meta-analysis suggest that retrieving the category name given an associated exemplar may be sufficient under some conditions to cause the forgetting of other exemplars from that category.

Next we examined the strength-independence assumption of RIF by analyzing the relationship between the enhancement of Rp+ items and the forgetting of Rp- items. A meta-regression analysis ( $k = 173$ ) indicated a significant positive association between the (standardized) memory enhancement effect of Rp+ items and memory inhibition effect of Rp- items (i.e., RIF effect),  $B = 0.19$ , 95% CI = [0.11, 0.26]. At the first sight, this seems to contradict the strength-independent assumption and thus provide evidence in support of competition-based accounts of RIF. However, it is possible that this association is limited to studies that failed to control output interference at test. If output interference does play a role in producing a given RIF effect, then the degree to which Rp+ items are strengthened should determine the degree to which Rp- items are forgotten. Note that this prediction can be made regardless of whether output interference is caused by competition and/or inhibition-based processes (see Anderson, 2003, for discussion).

To address this possibility, we conducted another meta-regression analysis limiting the analysis to samples that controlled output interference ( $k = 52$ ). Contrary to the above results, a significant relationship between the facilitation of Rp+ items and the forgetting of Rp- items was not observed (see Figure 2A),  $B = 0.02$ , 95% CI = [-0.12, 0.16]. On the other hand, when we limited our meta-regression analysis to studies that did not control output interference ( $k = 125$ ), the positive relationship between the facilitation of Rp+ items and the forgetting of Rp- items remained significant,  $B = 0.16$ , 95% CI = [0.06, 0.25] (see Figure 2B). This result indicates that the positive association can likely be explained by between-sample differences in output interference, and once output interference is controlled, there is no noticeable association.

Table 4 reports the average effect size for the samples that provided feedback during retrieval practice ( $k = 43$ ). The RIF effect ranged from small to medium and was statistically significant,  $g = 0.29$ , 95% CI = [0.19, 0.31]. However, we found no evidence that feedback strengthened the RIF effect. Indeed, the effect size was lower, though not significantly, than the overall effect size reported in Table 1,  $Q(1) = 1.36$ , ns. For completeness, we also analyzed the effect sizes as a function of whether studies controlled output interference. Consistent with the previous results, studies that did not control output interference exhibited a larger RIF effect,  $k = 18$ ,  $g = 0.51$ , 95% CI = [0.36, 0.66], than those that did control output interference,  $k = 27$ ,  $g = 0.18$ , 95% CI = [0.07, 0.30],  $Q(1) = 10.7$ ,  $p < .01$ .

Table 4 also reports the pooled effect size for samples that utilized the impossible retrieval-practice paradigm in which participants attempted to retrieve exemplars from retrieval cues that were not associated with any exemplars ( $k = 11$ ). Note that this paradigm does not have Rp+ items and therefore output interference is controlled by design; subjects are always directed to recall studied items at test, which includes only Rp- and Nrp items. The RIF effect with this paradigm was statistically significant,  $g = 0.28$ , 95% CI = [0.15, 0.40], and the effect size was comparable to the overall RIF effect for the studies that controlled output interference (see Table 1),  $Q(1) = 0.15$ , ns.

Lastly, we examined the relationship between RIF and performance during retrieval practice. The results were similar to those concerning the relationship between the enhancement of Rp+ items and the forgetting of Rp- items examined above. A meta-regression analysis indicated a significant positive association ( $k = 284$ ),  $B = 0.48$ , 95% CI = [0.20, 0.75] such that

studies involving higher rates of retrieval practice success reported greater levels of RIF than studies involving lower rates of retrieval practice success. When we limited the analysis to studies that controlled output interference ( $k = 153$ ), this association became non-significant,  $B = 0.14$ , 95% CI = [-0.22, 0.50]. A significant association was observed, however, when we limited the analysis to studies that did not control output interference ( $k = 147$ ),  $B = 0.48$ , 95% CI = [0.03, 0.94]. These results suggest that retrieval-practice success can predict larger RIF effects, but only when output interference is not controlled.

**Interference dependence.** Table 5 reports the RIF effect size for samples that directly manipulated the taxonomic frequencies of the exemplars ( $k = 12$ ). Both strong and weak exemplar conditions showed statistically significant RIF effects:  $g = 0.51$ , 95% CI = [0.35, 0.68] for the strong exemplar condition;  $g = 0.18$ , 95% CI = [0.03, 0.34] for the weak exemplar condition. Although we cannot conduct a mixed-effects model to test the difference between these effect sizes because the vast majority of the samples manipulated taxonomic frequency as a within-subject factor, the effect size is noticeably larger in the strong exemplar condition than in the weak exemplar condition. Indeed, the low end of the CI for strong exemplars was numerically higher than the high end of the CI for weak exemplars.

After excluding samples that did not control output interference ( $k = 4$ ), the RIF effect in the weak exemplar condition was no longer significant,  $g = 0.00$ , 95% CI = [-0.37, 0.39] whereas the RIF effect in the strong exemplar condition remained significant with a medium effect size,  $g = 0.49$ , 95% CI = [0.20, 0.78]. These results indicate that the small RIF effect in the weak exemplar condition can likely be attributed to dynamics at test, and not to inhibition during retrieval practice (see Anderson et al., 1994). In fact, the difference in the effect size between the samples that controlled output interference and those that did not was statistically significant,  $Q(1) = 4.27$ ,  $p < .05$ . Taken together, these findings provide support for the interference-dependence hypothesis of the inhibition-based account.

We observed a robust RIF effect when we examined the RIF effect in studies that used an episodic learning paradigm in which participants learned the materials in the context of the experiment ( $k = 175$ ),  $g = 0.33$ , 95% CI = [0.27, 0.39]. The effect remained significant even in studies controlling output interference ( $k = 99$ ),  $g = 0.21$ , 95% CI = [0.14, 0.28]. A mixed-effects model indicated that the effect size observed in these studies was not significantly different from that of studies that did not use an episodic learning paradigm,  $Q(1) = 1.07$ , ns. These findings might be interpreted as challenging inhibition-based accounts because ad-hoc, or episodically defined, categories do not have strong pre-existing associations, and items might therefore be less likely to interfere during retrieval practice.

**Individual differences.** The RIF effects in older adults, younger children, clinical patients (with deficits in executive functioning), and their control participants are described in Table 6. There are several studies that did not include a control condition (e.g., younger adults) to examine RIF in older adults or younger children (e.g., Conroy & Salmon, 2006). These studies were excluded from Table 6 to provide a fair comparison. For completeness, however, the analysis reported in the text will include samples without a control group.

Older adults ( $k = 13$ ) exhibited a statistically significant small-to-medium RIF effect,  $g = 0.42$ , 95% CI = [0.27, 0.57], indicating that RIF is a robust phenomenon even among older adults.

Although the effect size was numerically smaller than that of young adult controls ( $k = 13$ ),  $g = 0.47$ , 95% CI = [0.30, 0.64], the difference was not statistically significant  $Q(1) = 0.19$ , ns. A supplementary meta-analysis that added the older adult samples without a control condition ( $k = 17$ ) showed the same effect size estimate,  $g = 0.42$ , 95% CI = [0.27, 0.57]. Controlling output interference reduced the RIF effect for both older adults and young adult controls, but the RIF effect was statistically significant for both groups with small to medium effect sizes:  $g = 0.35$ , 95% CI = [0.20, 0.50] for older adults;  $g = 0.42$ , 95% CI = [0.24, 0.60] for younger adults. Once again, the difference between groups was not statistically significant,  $Q(1) = 0.38$ , ns.

Young children ( $k = 10$ ) exhibited a similar pattern. They showed a statistically significant RIF effect,  $g = 0.56$ , 95% CI = [0.37, 0.74], indicating that RIF is observed in young children. Although the effect size was numerically smaller than that of adult controls ( $k = 6$ ),  $g = 0.64$ , 95% CI = [0.37, 0.92], the difference was not statistically significant  $Q(1) = 0.27$ , ns. A supplementary meta-analysis that added the young children samples without control conditions ( $k = 20$ ) showed almost the same effect size estimate,  $g = 0.60$ , 95% CI = [0.46, 0.74]. Again, controlling output interference reduced the RIF effect for both young children ( $k = 3$ ) and adult controls ( $k = 2$ ), but the RIF effect sizes remained small to medium in both conditions:  $g = 0.41$ , 95% CI = [-0.08, 0.91] for young children;  $g = 0.57$ , 95% CI = [0.21, 0.93] for adult controls.

Clinical patients with presumed deficits in executive functioning also showed a significant RIF effect ( $k = 14$ ),  $g = 0.41$ , 95% CI = [0.15, 0.66]. The effect size was not statistically different from healthy control subjects ( $k = 17$ ),  $Q(1) = 1.55$ , ns, though healthy control subjects did show a numerically larger effect size,  $g = 0.62$ , 95% CI = [0.41, 0.83]. Unlike what we observed in the analyses of older adults and young children, however, when limiting the analysis to studies controlling output interference, the samples with deficits in executive functioning ( $k = 5$ ) failed to exhibit a significant RIF effect,  $g = -0.03$ , 95% CI = [-0.26, 0.20]. The control group ( $k = 5$ ) exhibited a numerically larger, but nonsignificant RIF effect,  $g = 0.22$ , 95% CI = [-0.10, 0.54], with the difference between the groups failing to reach significance,  $Q(1) = 1.55$ , ns.

### Other moderator analysis

Several additional moderator analyses are summarized in Table 7. We first examined the effects of final test type, focusing first on category-cued recall vs. category-plus-stem-cued recall. The results indicated that category-cued recall ( $k = 230$ ) showed a larger RIF effect,  $g = 0.47$ , 95% CI = [0.42, 0.51] than category-plus-stem cued recall ( $k = 112$ ),  $g = 0.27$ , 95% CI = [0.21, 0.33]. Although the RIF effect was significant in both tasks ( $ps < .01$ ), the difference between them was statistically significant,  $Q(1) = 26.4$ ,  $p < .01$ .

It should be noted that although studies using category-cued recall do not, for the most part, control output interference, some studies using an extralist paradigm (also called a semantic generation paradigm) were able to control output interference by having participants only recall Rp- and Nrp items at the time of the final test. Samples using this type of extralist paradigm ( $k = 28$ ) showed a significant RIF effect,  $g = 0.29$ , 95% CI = [0.21, 0.37], and the size of the effect was comparable to (or numerically larger than) the overall RIF effect reported in Table 1. We also compared the RIF effect sizes when the extralist paradigm employed a category-cued recall final test ( $k = 6$ ) and a category-plus-stem-cued recall final test ( $k = 18$ ). As observed in Table 7,

these two test types showed similar effect sizes:  $g = 0.30$ , 95% CI = [0.18, 0.41] for category-cued tests,  $g = 0.28$ , 95% CI = [0.17, 0.39] for category-plus-stem cued-recall tests, with a mixed-effects model failing to show a significant difference,  $Q(1) = 0.05$ , ns. This observation is remarkable, as it suggests that when output interference is controlled with the use of an extralist retrieval-practice paradigm, RIF no longer varies as a function of whether item-specific or category-general cues are employed at final test (see also our findings on independent probe paradigm).

Studies employing implicit memory tasks ( $k = 42$ ) showed a small, but statistically significant RIF effect,  $g = 0.18$ , 95% CI = [0.07, 0.28]. The effect remained significant, though somewhat reduced, when we focused on samples that controlled output interference ( $k = 34$ ),  $g = 0.12$ , 95% CI = [0.02, 0.22], or utilized a reaction time measure ( $k = 24$ ),  $g = 0.18$ , 95% CI = [0.07, 0.28]. Finally, the average RIF effect in studies using item-recognition tests was statistically significant ( $k = 58$ ),  $g = 0.32$ , 95% CI = [0.25, 0.39]. Given that recognition memory tasks are essentially item specific and therefore naturally control output interference (as long as Rp+ items are not deliberately tested first)<sup>4</sup>, we can expect that the effect size would be comparable with the overall RIF effect size that controlled output interference (see Table 1). Interestingly, the RIF effect observed using recognition memory tasks was significantly larger than that observed overall,  $Q(1) = 6.1$ ,  $p < .05$ .

We also investigated the relationship between the enhancement of Rp+ items and the forgetting of Rp- items for the studies specifically employing item-recognition final tests. A meta-regression analysis ( $k = 13$ ) failed to indicate a significant association between the (standardized) memory enhancement effect of Rp+ items and the forgetting effect of Rp- items,  $B = -0.08$ , 95% CI = [-0.36, 0.20]. This result is not consistent with the competition-based accounts of RIF observed using item-recognition final tests, as the extent to which Rp+ items were strengthened did not predict the amount of RIF that was observed. Given the small sample size, however, this finding should be interpreted with considerable caution.

When we examined RIF as a function of experimental design, between-subject designs ( $k = 37$ ) tended to show larger RIF effects,  $g = 0.51$ , 95% CI = [0.31, 0.70], than within-subject designs ( $k = 479$ ),  $g = 0.34$ , 95% CI = [0.31, 0.37]. The difference was marginally significant,  $Q(1) = 2.71$ ,  $p < 0.1$ . This finding is consistent with the idea that Nrp items may be deflated somewhat in the context of the typical within-subjects paradigm for assessing RIF (Tsukimoto & Kawaguchi, 2006). Specifically, participants may be inhibiting Nrp items as well as Rp- items in within-subject designs, thus making within-subject designs somewhat conservative in measuring RIF.

A meta-regression analysis was conducted to investigate the relationship between study time (in seconds) and RIF. As study time may be seriously confounded by type of material, with certain materials taking more time to study than others, we focused only on studies that used category-exemplar word pairs. A significant negative relationship was observed between study time and RIF for studies that controlled output interference ( $k = 153$ ),  $B = -0.03$ , 95% CI = [-0.05, -0.00], as well as for studies that did not control output interference ( $k = 137$ ),  $B = -0.04$ , 95% CI = [-0.08, -0.00]. Thus, longer study times appear to reduce the magnitude of RIF that is observed.

We also conducted a meta-regression analysis on the relationship between retrieval practice time (in seconds) and RIF. Again, we focused only on studies that used category-exemplar word pairs. A significant negative relationship was observed between retrieval practice time and RIF for studies that controlled output interference ( $k = 133$ ),  $B = -0.03$ , 95% CI =  $[-0.05, -0.01]$ , indicating that longer practice times can lead to smaller RIF effects. A significant relationship was not observed, however, for studies that did not control output interference ( $k = 103$ ),  $B = -0.01$ , 95% CI =  $[-0.05, 0.02]$ .

Finally, an analysis of the delay between retrieval practice and final test revealed that samples tested 24 hours or more after retrieval practice ( $k = 38$ ) failed to exhibit a statistically significant RIF effect,  $g = 0.09$ , 95% CI =  $[-0.07, 0.25]$ . This weak effect size was significantly different from the samples that performed a test session on the same day,  $g = 0.37$ , 95% CI =  $[0.34, 0.40]$ ,  $Q(1) = 19.2$ ,  $p < .01$ , suggesting that the RIF effect may diminish over time. However, a closer inspection also indicated that 45% of the samples that performed a delayed memory test used text materials. As described earlier, several studies using text materials were designed specifically to observe evidence of retrieval-induced facilitation (e.g., Chan, 2009). These studies may have distorted the effect size in the delayed samples. Accordingly, we did the same analysis with the samples that did not use text materials. This analysis provided a completely different picture about the effects of delay on RIF. Without text materials, the RIF effect after a delay ( $k = 21$ ),  $g = 0.40$ , 95% CI =  $[0.20, 0.61]$  was as large as the RIF effect without a delay ( $k = 371$ ),  $g = 0.35$ , 95% CI =  $[0.32, 0.38]$ . The small difference was not statistically significant,  $Q(1) = 0.01$ , ns. We further examined the delayed RIF effect by further dividing the samples into the papers that tested delayed RIF with the same set of items (i.e., retest;  $k = 6$ ) and those with a different set of items (i.e., no retest;  $k = 15$ ). The results showed that the RIF effect after delay was statistically significant regardless of whether the RIF was examined with retesting,  $g = 0.55$ , 95% CI =  $[0.08, 1.01]$  or without retesting,  $g = 0.35$ , 95% CI =  $[0.12, 0.59]$ . The difference was not statistically significant,  $Q(1) = 0.52$ , ns.

## General Discussion

The current research presented a comprehensive meta-analysis of the empirical studies on RIF. Across all studies, the analysis yielded a significant overall RIF effect (i.e., Rp- items were remembered less well than were Nrp items) that was observed across a broad range of study materials and experimental procedures. The moderator analyses, however, revealed a more nuanced picture, with RIF being subject to a number of boundary conditions and moderating factors. These moderator analyses largely supported inhibition-based theories of RIF, but at the same time posed some challenges. Below, we discuss some of the highlights of our findings, focusing in particular on those with theoretical implications.

### Output Interference

One factor that may contribute to retrieval-induced forgetting—at least partially, and under certain circumstances—is the way in which items are recalled at final test. Output interference is a form of RIF that can take place at the time of test, specifically when a subset of items is recalled before another set of items (Roediger, 1974; A. D. Smith, 1973; Tulving & Arbuckle, 1966). In the context of the retrieval-practice paradigm, because Rp+ items are strengthened by retrieval practice, they are likely to be recalled first on the final test, thus causing Rp- items to suffer output interference. There are ways to control for output interference, such as

by providing item-specific cues that determine the order in which items are recalled, or by testing Rp- items (and a baseline set of Nrp items) in a way that prevents Rp+ items from being recalled first.

In our meta-analysis, we compared the amount of RIF observed in studies that controlled for output interference versus those that did not. Across a variety of experimental materials and paradigms, controlling for output interference reduced the magnitude of the overall RIF effect, often by as much as cutting it in half. This observation is important for several reasons, but foremost because it suggests that a significant portion of the RIF that has been reported in the literature can be attributed to dynamics at test, rather than to the persisting consequence of dynamics that occurred during retrieval practice. However, the fact that robust levels of RIF were observed in studies that did control for output interference suggests that RIF cannot be explained entirely on the basis of output dynamics occurring at test.

As we discussed in the introduction, whether a study controls for output interference can be critical for testing theoretical accounts of RIF. Presumably, whatever inhibition mechanism does take place during retrieval practice—if one does indeed take place—the consequences of such a mechanism are going to be incurred regardless of how the final test is administered. If output interference is uncontrolled, however, then the contribution of such inhibition to the final forgetting effect can be obfuscated. Thus, failing to control for output interference can introduce a source of forgetting that makes it difficult to validly test certain theoretical predictions. For example, although forgetting caused by inhibition during retrieval practice may be strength independent or interference dependent, forgetting caused by output interference at test may mask such evidence, especially if a given effect of output interference is driven primarily by competition-based mechanisms.

### **Cue Independence**

The results of the meta-analysis provide some support for the cue-independence property of RIF. Specifically, independent cuing paradigms (i.e., the cross-category and extralist-cuing paradigms; Anderson & Spellman, 1995) were found to exhibit small, but statistically significant, effects of RIF. The finding that forgetting can be cue independent has served as one of the centerpieces in arguments supporting Anderson's version of the inhibitory account. If forgetting generalizes to independent cues, which it appears it can, then such evidence is difficult for competition-based accounts to explain. Although some have questioned the reliability of evidence for cue-independent RIF (Jonker et al., 2013; Perfect et al., 2004; Raaijmakers & Jakab, 2013), the results of the current meta-analysis suggest that while the effect may be smaller than the typical RIF effect, it is reliable. We also observed similar amounts of RIF using item-specific cues and non-item-specific cues (Table 2). This observation should be reassuring to those researchers who have argued that cue-independent designs are less susceptible to dynamics at test. If the size of the effect was different in the two types of studies, as it was for studies that did not use cue-independent designs, then one could argue that some test strategy was playing a role in producing the cue-independent RIF effect (see Camp et al., 2007). These findings further strengthen the theoretical implication of the findings from independent cuing paradigms.

The reason cue-independent paradigms exhibit smaller RIF effects than non-independent paradigms is less clear. One possibility is that the RIF observed using non-independent designs reflects the combined influence of a number of different factors. Cue-independent designs may lead to smaller forgetting effects because they effectively eliminate sources of forgetting that occur at test, or other non-inhibitory cue-dependent mechanisms, thus providing a more pure

assessment of the consequences of inhibition that occurred during retrieval practice. Another possibility is that inhibition can occur at both the level of an item's representation as well as at the level of the cue-target association, with the latter action making it particularly difficult to retrieve Rp- items in response to the original study cues. The inhibition of the cue-target association becomes irrelevant when independent cues are employed, however, thus reducing the amount of forgetting that is observed. Both of these possibilities provide reasonable explanations for the relatively reduced magnitude of the cue-independent RIF effect.

The independent cuing paradigm, however, has been criticized as a criterion to support the cue-independence property. According to some researchers, for example, the basic procedure proposed by Anderson and Spellman (1995) does not eliminate the possibility of covert cuing (Camp, Pecher, & Schmidt, 2005, 2007; Perfect et al., 2004). Specifically, when participants are given an independent cue at test (e.g., *monkey*, when participants study the pairs *fruit:orange* and *fruit:banana*), participants may be tempted to use the original category cue (i.e., *fruit*) to retrieve the items, making the independent cuing paradigm cue dependent. It should be worth noting, however, that covert cuing is not necessarily expected to produce forgetting, as covert cuing provides participants with multiple cues. Anderson (2003) argued that the memory benefit of multiple cues should be greater for Rp- items than for Nrp items, predicting that covert-cuing should eliminate, rather than boost the RIF effect (masking hypothesis). In support of this hypothesis, Weller, Anderson, Gómez-Ariza, and Bajo (2013) found that explicitly instructing participants to engage in covert cuing during a cue-independent final test eliminated the RIF effect (see also Huddleston & Anderson, 2012), a finding that poses a challenge to the idea that cue-independent RIF is the consequence of covert cuing. Moreover, these results suggest an alternative explanation for why RIF effects may be relatively reduced in cue-independent paradigms. Specifically, participants in many of the studies may have engaged in covert cuing, which may have in turn reduced the RIF effect observed in the present meta-analysis.

Finally, some proponents of inhibitory accounts have argued that item-recognition tests provide evidence of cue-independent forgetting, the logic being that the item cues used on recognition tests provide sufficient information to prevent competition from items that were strengthened by retrieval practice. If this rationale is correct, then the robust level of RIF observed in studies employing item-recognition tests provides additional evidence that RIF is cue independent. This rationale, however, has been questioned. Verde and Perfect (2011), for example, argued that the fact that RIF is observed on item-recognition tests does not necessarily provide support for cue independence. They argued that recognition memory performance entails a recollection component, which necessitates a context-dependent search for episodic details (Yonelinas, 2002). Importantly, recollection allows participants to utilize a variety of cues (including category cues during study or retrieval practice) to identify the target. Verde and Perfect found evidence that it is this recollection component that causes RIF in recognition memory, indicating that the recognition memory task may not provide a valid test of cue independence, (but see Spitzer & Bäuml, 2007; and our meta-regression results, which failed to find a significant relationship between strengthening and RIF in studies using item-recognition final tests). Given the robust levels of RIF observed in recognition tests in the current meta-analysis, future research should focus more on how and why RIF is observed on recognition tests, and less on whether it is observed on recognition tests.

### **Retrieval Specificity and Strength Independence**

The retrieval specificity property received mixed support in the meta-analysis. First, in

support of the property, no evidence of RIF was observed in studies involving the restudying of a subset of items during what would normally be retrieval practice. This finding is difficult for competition-based theories to explain because strengthening a subset of items—even if through restudy—should have been sufficient to cause forgetting. This finding is consistent with the retrieval-specificity assumption of some inhibition-based accounts, however, in that simply restudying items may not require the involvement of inhibition because other items presumably do not cause interference during restudy. Interestingly, a very different pattern of results was observed in studies employing the category-retrieval form of practice (i.e., *Fr\_\_\_\_\_Banana*). In contrast to the frequently cited findings by Anderson et al. (2000), retrieving category names given their associated exemplars as retrieval cues does appear to be sufficient to cause RIF (Table 3). This observation is consistent with competition-based accounts of RIF and suggests that RIF may not be retrieval specific.

There are several possible explanations for why studies using the restudy and category-retrieval paradigms exhibited such different patterns of results. Competition-based accounts posit that strengthening the association between the categories and exemplars should yield additional interference, regardless of whether the association is strengthened via retrieval of the exemplar or of the category. Thus, one possibility is that retrieving the categories strengthened practiced items to a greater extent than simple restudy, causing strength-based competition at the time of test (but see our findings on the null relationship between the strengthening caused by retrieval practice and RIF discussed below, which would seem to argue against this strength-based account). Another possibility, supported by inhibition-based accounts, is that category retrieval practice may actually require the resolution of competition, thus causing inhibition. This may especially be the case when category retrieval task is difficult. Finally, the discrepant findings may be also explained by the context-cuing account. Specifically, because the category-retrieval task may be more likely to cause a shift in context between the study phase and practice phase (see Jonker et al., 2013), it may make it more likely for participants to search the inappropriate practice phase for non-practiced items related to practiced categories at final test. Further research is needed to test these different accounts of our findings.

Other evidence relevant to strength independence is much more difficult for competition-based accounts to explain. For example, receiving feedback during retrieval practice failed to increase the RIF effect, and receiving impossible retrieval practice—or failing to perform well during retrieval practice—failed to decrease the RIF effect (relative to the overall RIF effect). Moreover, as long as output interference was controlled at final test, RIF was not observed in studies employing exemplars of low-taxonomic frequency. Each of these findings provides evidence that strengthening  $R_{p+}$  items is insufficient to cause RIF. Although there have been failures to replicate (e.g., Williams & Zacks, 2001), and individual studies have sometimes provided contradicting evidence, the overall pattern of results observed in this meta-analysis clearly supports the strength-independence assumption.

One new line of evidence reported here is that as long as output interference is controlled at final test, the extent to which  $R_{p+}$  items are strengthened by retrieval practice does not predict the extent to which  $R_{p-}$  items are forgotten (Figure 2). This finding supports the strength independence property and is difficult for competition-based accounts to explain because studies involving greater strengthening of  $R_{p+}$  items should have led to larger forgetting effects. Interestingly, a significant positive correlation was observed in studies that did not control for output interference. This juxtaposition provides a nice demonstration of how the property of

strength independence may only apply to studies that adequately control for output interference at test. When output interference is not controlled, the extent to which Rp+ items are strengthened may indeed have a strong effect on whether Rp- items are forgotten. The fact that the association was not just reduced, but eliminated, when output interference was controlled, suggests that strength-based competition mechanisms may play very little role in causing forgetting in such situations.

### **Interference Dependence**

The primary evidence relevant to the property of interference dependence investigated in the present meta-analysis was that of taxonomic frequency. Specifically, Anderson et al. (1994) found that exemplars of high taxonomic strength suffered significantly more forgetting than did exemplars of relatively low taxonomic strength. Although based on a relatively small sample of studies, the results of the current meta-analysis suggest that this finding is reliable (see also Migueles & Garcia-Bajos, 2014, for a recent replication). Interestingly, whether exemplars of low taxonomic strength suffered RIF at all depended on whether output interference was controlled at final test: when output interference was not controlled, a significant effect was observed; when output interference was controlled, a significant effect was not observed. This pattern is consistent with the idea that although low frequency exemplars may not be susceptible to inhibition during retrieval practice, they may indeed be susceptible to output interference at final test.

It should be noted that we observed a robust RIF effect in studies that employed materials with ad-hoc, or episodically defined, categories. Some researchers have argued that the RIF effect with ad-hoc or episodically defined categories challenges the interference dependence property because such materials have no a priori associations and should thus function like “weak” category-exemplar pairs function in studies that manipulate taxonomic frequency. One factor that may help explain these seemingly conflicting lines of evidence is that ad-hoc items may be more likely to cause interference and be inhibited at the episodic level compared to category exemplars. That is, when attempting to retrieve ad-hoc items during retrieval practice, subjects might search episodically from the earlier study phase, leading to the inhibition and forgetting of other items from that episode. When attempting to retrieve category exemplars, however, subjects may be more likely to search semantic memory, and do so in a way that does not require low-frequency exemplars to be inhibited.

Although our meta-analysis provided some evidence concerning the reliability of interference dependence, it should be emphasized that taxonomic frequency is not the only way to empirically test this property. Interference dependence has also been tested in other ways that were not amenable to inclusion in our meta-analysis (due to a small number of studies). For example, Shivde and Anderson (2001) found that retrieval practice for the subordinate meaning of a homograph caused the forgetting of the dominant meaning, whereas retrieval practice for the dominant meaning failed to cause forgetting of the subordinate meaning. Because the dominant meaning is more likely to interfere with the retrieval of the subordinate meaning, but not vice versa, this finding has been interpreted as evidence for interference dependence. Another relevant finding is that a list of items that participants are instructed to remember is more susceptible to RIF than is a list of items that participants are instructed to forget (e.g., Bäuml & Sameniéh, 2010; Storm et al., 2007). According to inhibition-based theories, the instruction to remember a set of items makes those items more likely to suffer RIF because they are more likely to interfere with retrieval practice. Other research has provided evidence inconsistent with

interference dependence. For example, Jakab and Raaijmakers (2009) found that RIF did not vary as a function of study position. They reasoned that, according to the inhibition account, items studied early on in a list should have been more susceptible to RIF because they should have been more accessible during retrieval practice and thus more likely to interfere with retrieval practice.

Given the importance of the interference-dependence property for inhibitory-based accounts of RIF, it is surprising that there have not been more studies designed to test it. One reason for this paucity may be that it can be difficult to determine beforehand whether a given manipulation increases or decreases interference during retrieval practice. Study time, for example, might be thought of as a manipulation that increases interference by making studied items more likely to interfere. However, increasing study time is also likely to increase integration of studied items in a way that actually decreases interference (see also Anderson, Green, & McCulloch, 2000; Goodman & Anderson, 2011). The interference dependence assumption has also faced recent modification. Anderson and Levy (2011) argued that two opposing forces may come into play when a competitor becomes more interfering. First, increased interference should raise the demand for inhibition, increasing the likelihood that inhibition is triggered. Second, increased interference should decrease the probability that inhibition will successfully deactivate the opposing representation. As a consequence of these conflicting processes (referred to as the demand-success trade-off), Anderson and Levy (2011) indicated that the relationship between interference and memory impairment may not reflect a purely monotonic relationship (see Keresztes & Racsmany, 2013, for empirical support). Given these complexities, future research will likely need to develop new experimental paradigms that are more capable of testing the interference-dependence assumption in more nuanced and stringent ways.

### **Individual Differences**

The results of our meta-analysis suggest that young children and older adults exhibit RIF effects comparable to control participants (i.e., young adults, including undergraduate students), a finding that was observed even when the samples were limited to studies that controlled for output interference. Because young children and older adults have been argued to have deficiencies in executive functioning, these results seem to be inconsistent with the notion that RIF is caused by the top-down inhibitory control processes of executive functioning. Two notes should be made. First, one might argue that the older adults volunteering to participate in research studies do not reflect the average person of that age, and that their increased motivation to perform well on the task (relative to undergraduate subject pool participants) makes them more likely to exhibit large effects of RIF even if general executive functions do become diminished with age. This explanation, however, cannot easily account for the lack of impairment in young children. Second, Aslan and Bäuml (2010, 2012) showed the absence of RIF for younger children and older adults, but only when they focused on very young children (e.g., kindergartners) or very old adults (old-old adults; 75-95 years old). These findings indicate that the RIF effect observed in older adults and young children in the literature may have been caused by the fact that previous research sampled relatively mature younger children (e.g., schoolchildren) or relatively young older adults. Moreover, although Ortega, Gómez-Ariza, Román, and Bajo (2012) found that young older adults exhibited significant RIF, this effect was eliminated when the young older adults engaged in the retrieval-practice task while also engaging in a moderately demanding secondary task. Importantly, the same moderately

demanding task did not eliminate RIF for younger adults. This observation indicates that young older adults may not be old enough to show a clear deficit in RIF, but that they may be on the edge of doing so.

The results of the meta-analysis also suggest that clinical patients who exhibit deficits in executive functioning exhibit impaired levels of RIF, but only in studies that control for output interference at test. However, control condition participants also failed to exhibit significant RIF (though they indicated a numerically larger RIF effect), making it unclear whether the non-significant RIF effect can be attributed to deficits in executive functioning of patients or idiosyncratic experimental characteristics of the studies included in the analysis. These findings, together with the findings with younger children and older adults, pose some challenge to inhibition-based accounts. As the number of studies included in the meta-analysis is relatively small (especially the studies that controlled output interference), future work is clearly needed to examine the robustness and generality of these findings.

Researchers have recently moved beyond special populations to examine other individual differences related to RIF, with each study typically computing a RIF score for each subject and then examining the correlation between those RIF scores and some other index, such as working memory capacity, autobiographical memory, fixation in problem solving, and so forth (e.g., Aslan & Bäuml, 2011; Koppel & Storm, 2014; Lev-Ari & Peperkamp, 2013; Schilling, Storm, & Anderson, 2014; Storm & Angello, 2010; Storm & Jobe, 2012a). The idea is that if RIF is primarily the consequence of inhibition, then individuals who exhibit more RIF should exhibit the types of advantages associated with inhibition (for more nuanced discussions of individual differences and RIF, see Anderson & Levy, 2007; 2011). If RIF is largely the consequence of competition at test, however, then we should not expect to observe these sorts of correlations. In fact, according to most competition-based accounts, individuals who exhibit more RIF should be, if anything, more prone to blocking and interference. Interestingly, most individual difference studies to date have supported the inhibition-based account, finding significant positive correlations between RIF and the types of advantages associated with inhibition.

### **Design Parameters**

A number of task variables and parameters related to the retrieval-practice paradigm were investigated in the current meta-analysis. Below we briefly review the main findings associated with each phase of the paradigm.

**Study Phase.** A significant association was observed between RIF and the amount of time participants were exposed to items in the study phase. Interestingly, the association was negative such that items exposed for longer durations tended to suffer less RIF than items exposed for shorter durations. One interpretation of this finding is that longer exposure times may encourage subjects to use encoding strategies or integrate items in a way that prevents the items from being susceptible to RIF (e.g., Goodmon & Anderson, 2011). Researchers hoping to observe large RIF effects in the future should consider employing relatively short study trials.

**Retrieval Practice Phase.** A similar association was observed between RIF and the amount of time participants were given to retrieve items during retrieval practice. Specifically, studies giving participants less time to respond on retrieval practice trials exhibited larger forgetting effects than those giving participants more time to respond. The explanation for this association is not immediately obvious. One possibility is that giving participants too much time allows them to engage in covert rehearsal for non-practiced items. For example, after successfully retrieving the targeted *fruit* they might spend the rest of the time thinking about the

other *fruits* they had studied. If this covert rehearsal was biased toward practiced categories, and it seems likely that it would be, then it might account for the reduction in RIF observed on the final test (for a similar discussion, see Goodmon & Anderson, 2011).

Another interesting finding is the relative unimportance of performance during retrieval practice. In studies controlling for output interference, the extent to which participants performed well during retrieval practice did not predict the extent to which participants exhibited RIF on the final test. This finding fits well with the finding in our meta-analysis that impossible retrieval practice causes just as much forgetting as possible retrieval practice (Table 4). It seems that RIF is decidedly not contingent on retrieval-practice success. Moreover, it also seems unnecessary to provide corrective feedback during retrieval practice, as we found no evidence that RIF was increased in studies that provided such feedback.

**Semantic Generation/Extra-List Retrieval Practice.** A number of studies have employed a modified version of the retrieval-practice paradigm in which participants generate and then repeatedly retrieve new exemplars associated with a subset of categories during retrieval practice, as opposed to retrieving a subset of items from the initial study phase (e.g., Bäuml, 2002; Storm et al., 2008). The current meta-analysis suggests that studies using this form of semantic generation, or extra-list retrieval practice, exhibit just as much RIF as do studies that use the more typical form of episodic retrieval practice. One advantage of the semantic generation methodology is that participants can be asked to recall the studied items directly without asking them to also recall the items generated during retrieval practice, thus providing for a natural and perhaps more effective control of output interference. Indeed, as would be expected if output interference was controlled, unlike with the typical paradigm, the amount of RIF observed following semantic generation retrieval practice did not differ between studies employing category-plus-stem final test cues and studies employing category final test cues.

**Retention Interval.** A finding that may be surprising to many researchers in the field is that RIF can be observed in studies using long retention intervals, even those lasting as long as 24 hours. An often cited study by M. D. MacLeod and Macrae (2001) failed to find evidence of a long-lasting RIF effect, but the results of our meta-analysis suggest that subsequent evidence has overturned this observation. Because long-term memory performance involves a variety of mechanisms such as memory consolidation (McGaugh, 2000), and involuntary recollection (Conway & Pleydell-Pearce, 2000), it can be difficult to draw firm theoretical implications from this finding. Moreover, some of the studies tested long-term RIF using a test-retest design in which the same items were tested multiple times, which can pose interpretative difficulties because it is possible that the robust long-term RIF effects can be attributed to the fact that participants had the benefit of recalling more *Nrp* items than *Rp-* at a short delay. The results of our meta-analysis, however, suggest that significant RIF can be observed after long delays even when test-retest designs are not employed (e.g., Garcia-Bajos et al., 2009; Storm et al., 2012). In any case, the fact that RIF can have persisting consequences should be considered when interpreting the practical real-world implications of the phenomenon.

**Final Test.** A number of final test formats have been used in the study of RIF, including: category-cued recall, category-plus-stem-cued recall, independent-cued recall, item recognition, implicit memory tasks, and tasks assessing reaction time. Each format was found to produce reliable RIF effects, though the magnitude of the effects did differ between the different formats. Category-cued recall led to the largest forgetting effects, whereas independent-cued recall and implicit memory tasks led to the smallest forgetting effects. Whether these different methods of

assessing RIF reflect differences in sensitivity to a common set of mechanisms or differences in sensitivity to distinct sets of mechanisms, however, remains to be seen.

### **Limitations of the Meta-Analysis**

Although we believe that the results of this meta-analysis are informative, and that they have great potential for directing future research in more constructive directions, they should also be interpreted with great caution. Meta-analysis provides a reasonable way to summarize and quantitatively synthesize a large number of previous empirical studies, but it cannot account for the idiosyncratic design features of individual experiments (Lipsey & Wilson, 2001). Thus, the results may be biased by systematic confounding factors that correlate with effect size. Stated differently, one limitation of all meta-analysis findings is that they are correlational, not experimental. This issue may be especially problematic for moderation analyses that include only small numbers of studies, such as those concerning the restudy paradigm, manipulation of taxonomic frequency, and clinical patients with executive control deficits. Relatedly, although the results from meta-regression analyses provide some useful insights into the quantitative relationships between some independent variables (e.g., Rp+ performance) and RIF, these relationships are based on between-study comparisons, and the methodological literature indicates that care must be taken when attempting to generalize correlations based on between-participant or between-study variation into within-person psychological mechanisms (Hamaker, Dolan, & Molenaar, 2005; Molenaar & Campbell, 2009).<sup>5</sup> Thus, although our meta-analysis provides a number of theoretical implications, we should refrain from drawing overly strong inferences.

### **Concluding Comment**

The present meta-analysis provides an assessment of the literature that would be difficult, if not impossible, to achieve using a qualitative approach. It has become common practice, when reviewing research on RIF, for authors to cite evidence that supports their particular perspective while ignoring or downplaying evidence refuting it. Although not ideal, this practice is understandable. The literature has become so large that it is often impossible to adequately represent the totality of the evidence, especially when authors can persuade themselves—rightly or wrongly—that certain evidence is more convincing than other evidence and that certain methodologies are more appropriate than other methodologies. The advantage of the current meta-analysis is that it provides a more objective snap-shot of the literature, one that collapses across most of the studies that have been conducted.

By providing a more objective and comprehensive snap-shot, we hope the current meta-analysis will foster new and more effective directions for future research. One step, in particular, will be to further develop the nature and predictions of the various theoretical accounts (see Anderson & Bjork, 1994, for one example). Clearly there are different ways in which inhibition can be instantiated, yet by-in-large the focus of the literature has been on distinguishing between inhibitory and non-inhibitory accounts, as opposed to distinguishing between different instantiations of these accounts. Along these lines, the theoretical properties outlined by Anderson (2003)—cue independence, retrieval specificity, strength independence, and interference dependence—have been instrumental in work on RIF, and they continue to drive research today, but not all instantiations of the inhibition account require these properties to be true. Cue independence, for example, falls out of the argument that inhibition acts at the level of an item's representation, but one could easily imagine that inhibition acts at the level of the association between two items. Indeed, rendering interfering items less accessible in the context

in which they cause interference, while allowing them to maintain their accessibility in other contexts, would seem to be a very adaptive consequence of a purported inhibitory mechanism (e.g., Bjork, 1989; Bjork & Bjork, 1992). The broader point, however, is that as research on RIF continues to progress, it will be crucial for researchers to avoid overgeneralizing the theoretical implications of their findings, and to instead design their studies and interpret their results in relation to specific instantiations of a given theoretical account.

## References

(References marked with an asterisk indicate studies used in the meta-analysis)

- \* Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory*, *20*, 420-428. doi: 10.1080/09658211.2012.671832
- \* AhnAllen, C. G., Nestor, P. G., McCarley, R. W., & Shenton, M. E. (2007). The role of retrieval inhibition in the associative memory impairment of schizophrenia. *Psychiatry Research*, *150*, 43-50. doi: 10.1016/j.psychres.2006.07.001
- \* Amir, N., Badour, C. L., & Freese, B. (2009). The effect of retrieval on recall of information in individuals with posttraumatic stress disorder. *Journal of Anxiety Disorders*, *23*, 535-540. doi: 10.1016/j.janxdis.2008.10.012
- \* Amir, N., Coles, M. E., Brigidi, B., & Foa, E. B. (2001). The effect of practice on recall of emotional information in individuals with generalized social phobia. *Journal of Abnormal Psychology*, *110*, 76-82. doi: 10.1037/0021-843X.110.1.76
- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415-445. doi: 10.1016/j.jml.2003.08.006
- \* 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*, 544-570. doi: 10.1037/0096-3445.130.3.544
- \* Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review*, *7*, 522-530. doi: 10.3758/BF03214366
- Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265-325). New York: Academic Press.
- \* 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*, 1063-1087. doi: 10.1037/0278-7393.20.5.1063
- \* Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1141-1159. doi: 10.1037/0278-7393.26.5.1141
- Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory *Inhibition in cognition*. (pp. 81-102): American Psychological Association, Washington, DC

- Anderson, M. C., & Levy, B. J. (2011). On the relationship between interference and inhibition in cognition *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork*. (pp. 107-132): Psychology Press, New York, NY
- \* Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 608-629. doi: 10.1037/0278-7393.25.3.608
- \* Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68-100. doi: 10.1037/0033-295X.102.1.68
- \* Appan, R., & Browne, G. J. (2010). Investigating Retrieval-Induced Forgetting During Information Requirements Determination. *Journal of the Association for Information Systems*, *11*, 250-275.
- \* Aslan, A., Bäuml, K.-H. T., & Pastötter, B. (2007). No inhibitory deficit in older adults' episodic memory. *Psychological Science*, *18*, 72-78. doi: 10.1111/j.1467-9280.2007.01851.x
- \* Aslan, A., & Bäuml, K.-H. T. (2010). Retrieval-induced forgetting in young children. *Psychonomic Bulletin & Review*, *17*, 704-709. doi: 10.3758/PBR.17.5.704
- \* Aslan, A., & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 264-269. doi: 10.1037/a0021324
- \* Aslan, A., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting in old and very old age. *Psychology and Aging*, *27*, 1027-1032. doi: 10.1037/a0028379
- Bäuml, K.-H. T. (1998). Strong items get suppressed, weak items do not: the role of item strength in output interference. *Psychonomic Bulletin & Review*, *5*, 459-463.
- \* Bäuml, K.-H. T. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, *13*, 356-360. doi: 10.1111/j.0956-7976.2002.00464.x
- Bäuml, K.-H. T. (2007). Making memories unavailable: The inhibitory power of retrieval. *Zeitschrift für Psychologie/Journal of Psychology*, *215*, 4-11. doi: 10.1027/0044-3409.215.1.4
- \* Bäuml, K.-H. T., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition*, *32*, 610-617. doi: 10.3758/BF03195852
- \* Bäuml, K.-H. T., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory*, *10*, 215-224. doi: 10.1080/09658210143000362
- \* Bäuml, K.-H. T., & Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Memory & Cognition*, *31*, 1188-1197. doi: 10.3758/BF03195802

- \* Bäuml, K.-H., & Kuhbandner, C. (2007). Remembering can cause forgetting --- but not in negative moods. *Psychological Science*, *18*, 111-115. doi: 10.1111/j.1467-9280.2007.01857.x
- Bäuml, K.-H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science*, *21*, 793-795. doi: 10.1177/0956797610370162
- \* Bäuml, K.-H., Zellner, M., & Vilimek, R. (2005). When remembering causes forgetting: Retrieval-induced forgetting as recovery failure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1221-1234. doi: 10.1037/0278-7393.31.6.1221
- \* Bäuml, K.-H. T., & Samenieh, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 488-494. doi: 10.1037/a0025683
- \* Bajo, M. T., Gomez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-induced forgetting in perceptually driven memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1185-1194. doi: 10.1037/0278-7393.32.5.1185
- Baran, B., Wilson, J., & Spencer, R. M. C. (2010). Rem-dependent repair of competitive memory suppression. *Experimental Brain Research*, *203*, 471-477. doi: 10.1007/s00221-010-2242-2
- Barch, D. M. (2009). Neuropsychological abnormalities in schizophrenia and major mood disorders: Similarities and differences. *Current Psychiatry Reports*, *11*, 313-319. doi: 10.1007/s11920-009-0045-6
- Barkley, R. A. (1997). Attention-deficit/hyperactivity disorder, self-regulation, and time: Toward a more comprehensive theory. *Journal of Developmental and Behavioral Pediatrics*, *18*, 271-279. doi: 10.1097/00004703-199708000-00009
- \* Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion*, *18*, 457-477. doi: 10.1080/0269993034000392
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates
- 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*. Hillsdale, NJ: Erlbaum
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum

- Bjork, R. A., Bjork, E. L., & Caughey, B. J. (2007). Retrieval as a self-limiting process: Part II *The foundations of remembering: Essays in honor of Henry L. Roediger, III.* (pp. 19-37): Psychology Press, New York, NY.
- Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive-development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*, 48-71. doi: 10.1016/0273-2297(90)90004-n
- \* Blix, I., & Brennen, T. (2012). Retrieval-induced forgetting after trauma: A study with victims of sexual assault. *Cognition and Emotion, 26*, 321-331. doi: 10.1080/02699931.2011.570312
- Bond, C. F., Jr., Wiitala, W. L., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods, 8*, 406-418. doi: 10.1037/1082-989X.8.4.406
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). Comprehensive meta-analysis (Version 2.2.048) [Computer software].
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, United Kingdom: Wiley.
- \* Brown, A. D., Kramer, M. E., Romano, T. A., & Hirst, W. (2012). Forgetting trauma: Socially shared retrieval-induced forgetting and post-traumatic stress disorder. *Applied Cognitive Psychology, 26*, 24-34. doi: 10.1002/acp.1791
- \* Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1314-1319. doi: 10.1037/0278-7393.27.5.1314
- Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64*, 363-423. doi: 10.3102/00346543064003363
- \* Camp, G., Pecher, D., & Schmidt, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review, 12*, 490-494. doi: 10.3758/BF03193793
- \* Camp, G., Pecher, D., & Schmidt, H. G. (2007). No retrieval-induced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 950-958. doi: 10.1037/0278-7393.33.5.950
- \* Camp, G., Wesstein, H., & De Bruin, A. B. H. (2012). Can questioning induce forgetting? Retrieval-induced forgetting of eyewitness information. *Applied Cognitive Psychology, 26*, 431-435. doi: 10.1002/acp.2815
- Campbell, J. I. D., & Thompson, V. A. (2012). Retrieval-induced forgetting of arithmetic facts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 118-129. doi: 10.1037/a0025056

- \* Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology, 19*, 580-606. doi: 10.1080/09541440701326071
  - \* Caughey, J. E. (2008). *Inhibitory dynamics in human memory: From criminal interrogation to classroom instruction*. (Doctoral dissertation, AAI3302591).
  - \* Chan, J. C. K. (2008). *An investigation of retrieval-induced facilitation*. (Doctoral dissertation, AAI3299946).
  - \* Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61*, 153-170. doi: 10.1016/j.jml.2009.04.004
  - \* Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18*, 49-57. doi: 10.1080/09658210903405737
  - \* Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553-571. doi: 10.1037/0096-3445.135.4.553
  - \* Chiu, C.-D., Lin, C.-C., Yeh, Y.-Y., & Hwu, H.-G. (2012). Forgetting the unforgotten affective autobiographical memories in nonclinical dissociators. *Emotion, 12*, 1102-1110. doi: 10.1037/a0025900
  - \* Chiu, C.-D., Yeh, Y.-Y., Huang, C.-L., Wu, Y.-C., Chiu, Y.-C., & Lin, C.-C. (2010). Unintentional memory inhibition is weakened in non-clinical dissociators. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 117-124. doi: 10.1016/j.jbtep.2009.11.003
  - \* Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403-1414. doi: 10.1037/0278-7393.25.6.1403
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed. ed.). Hillsdale, NJ: Lawrence Erlbaum.
- \* Collette, F., Germain, S., Hogge, M. I., & Van der Linden, M. (2009). Inhibitory control of memory in normal ageing: Dissociation between impaired intentional and preserved unintentional processes. *Memory, 17*, 104-122. doi: 10.1080/09658210802574146
  - \* Coman, A., & Hirst, W. (2012). Cognition through a social network: The propagation of induced forgetting and practice effects. *Journal of Experimental Psychology: General, 141*, 321-336. doi: 10.1037/a0025247
  - \* Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science, 20*, 627-633. doi: 10.1111/j.1467-9280.2009.02343.x

- Conroy, R., & Salmon, K. (2005). Selective postevent review and children's memory for nonreviewed materials. *Journal of experimental child psychology*, *90*, 185-207.
- \* Conroy, R., & Salmon, K. (2006). Talking about parts of a past experience: The impact of discussion style and event structure on memory for discussed and nondiscussed information. *Journal of Experimental Child Psychology*, *95*, 278-297. doi: 10.1016/j.jecp.2006.06.001
- \* Conway, M. A., & Fthenaki, A. (2003). Disruption of inhibitory control of memory following lesions to the frontal and temporal lobes. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *39*, 667-686. doi: 10.1016/S0010-9452(08)70859-1
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*, 261. doi: 10.1037/0033-295X.107.2.261
- \* Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science*, *18*, 727-733. doi: 10.1111/j.1467-9280.2007.01967.x
- \* Dehli, L., & Brennen, T. (2009). Does retrieval-induced forgetting occur for emotional stimuli? *Cognition and Emotion*, *23*, 1056-1068. doi: 10.1080/02699930802285221
- \* Dodd, M. D., Castel, A. D., & Roberts, K. E. (2006). A strategy disruption component to retrieval-induced forgetting. *Memory & Cognition*, *34*, 102-111. doi: 10.3758/BF03193390
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170-177. doi: 10.1037/1082-989X.1.2.170
- \* Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, *39*, 420-433. doi: 10.1016/S0022-1031(03)00032-5
- \* Edgington, T., & Rusted, J. M. (2003). Separate and combined effects of scopolamine and nicotine on retrieval-induced forgetting. *Psychopharmacology*, *170*, 351-357.
- Enders, C. K. (2010). *Applied missing data analysis*: Guilford Press, New York, NY.
- \* Ford, R. M., Keating, S., & Patel, R. (2004). Retrieval-induced forgetting: A developmental study. *British Journal of Developmental Psychology*, *22*, 585-603. doi: 10.1348/0261510042378272
- Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory*, *17*, 92-103.

- Garcia-Bajos, E., & Migueles, M. (2013). Script-driven processing affords protection from retrieval-induced forgetting in the recall of everyday activities. *Quarterly Journal of Experimental Psychology*, *66*, 1317-1330. doi: 10.1080/17470218.2012.739184
- Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, *18*, 271-279. doi: 10.2307/1165136
- \* Gómez-Ariza, C. J., Fernandez, A., & Bajo, M. T. (2012). Incidental retrieval-induced forgetting of location information. *Psychonomic Bulletin & Review*, *19*, 483-489. doi: 10.3758/s13423-012-0228-8
- \* Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition*, *33*, 1431-1441. doi: 10.3758/BF03193376
- \* Gómez-Ariza, C. J., Pelegrina, S., Lechuga, M. T., Suárez, A., & Bajo, M. T. (2009). Inhibition and retrieval of facts in young and older adults. *Experimental Aging Research*, *35*, 83-97. doi: 10.1080/03610730802545234
- \* Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 416-436. doi: 10.1037/a0021963
- \* Groome, D., & Grant, N. (2005). Retrieval-induced forgetting is inversely related to everyday cognitive failures. *British Journal of Psychology*, *96*, 313-319. doi: 10.1348/000712605X49006
- \* Groome, D., & Sterkaj, F. (2010). Retrieval-induced forgetting and clinical depression. *Cognition and Emotion*, *24*, 63-70. doi: 10.1080/02699930802536219
- \* Groome, D., Thorne, J. D., Grant, N., & Pipilis, Y. J. (2008). Retrieval-induced forgetting and unwanted thought intrusions. *European Journal of Cognitive Psychology*, *20*, 723-737. doi: 10.1080/09541440701811965
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207-233. doi: 10.1207/s15327906mbr4002\_3
- \* Hanslmayr, S., Staudigi, T., Aslan, A., & Bäuml, K.-H. T. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive, Affective & Behavioral Neuroscience*, *10*, 329-338. doi: 10.3758/CABN.10.3.329
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193-225). New York: Academic Press

- \* Hauer, B. J. A., Wessel, I., Merckelbach, H., Roefs, A., & Dalgleish, T. (2007). Effects of repeated retrieval of central and peripheral details in complex emotional slides. *Memory, 15*, 435-449. doi: 10.1080/09658210701312085
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504. doi: 10.1037/1082-989X.3.4.486
- \* Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11*, 125-130. doi: 10.3758/BF03206471
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560. doi: 10.1136/bmj.327.7414.557
- \* Hogge, M. I., Adam, S. p., & Collette, F. (2008). Retrieval-induced forgetting in normal ageing. *Journal of Neuropsychology, 2*, 463-476. doi: 10.1348/174866407X268533
- 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*, 1408-1418. doi: 10.1037/a0027092
- \* Hughes, A. D. (2007). *Retrieval-induced forgetting: A case of interference*. (Doctoral dissertation, AAINR16954).
- \* Hulbert, J. C., Shivde, G., & Anderson, M. C. (2012). Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. *Experimental Psychology, 59*, 11-21. doi: 10.1027/1618-3169/a000120
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*: Sage Publications, Inc, Thousand Oaks, CA.
- \* Iglesias-Parro, S., & Gomez-Ariza, C. J. (2006). Biasing decision making by means of retrieval practice. *European Journal of Cognitive Psychology, 18*, 899-908. doi: 10.1080/09541440500501189
- \* Iglesias-Parro, S., Gomez-Ariza, C. J., & Arias, A. V. (2009). Inhibition as an adaptive mechanism in memory-based choices. *Revista de Psicología Social, 24*, 333-347. doi: 10.1174/021347409789050588
- \* Jakab, E. k., & Raaijmakers, J. G. W. (2009). The role of item strength in retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 607-617. doi: 10.1037/a0015264
- \* Johnson, S. K., & Anderson, M. C. (2004). The Role of Inhibitory Control in Forgetting Semantic Knowledge. *Psychological Science, 15*, 448-453. doi: 10.1111/j.0956-7976.2004.00700.x

- \* Johnston, L. J. (2005). *Examining the mechanism and influential extent of retrieval-induced forgetting*. (Doctoral dissertation, AAI3145080).
- \* Jonker, T. R., & MacLeod, C. M. (2012). Retrieval-induced forgetting: Testing the competition assumption of inhibition theory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale*, *66*, 204-211. doi: 10.1037/a0027277
- Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review*, *120*, 852-872.
- \* Kato, K. (2007). *Controlled versus automatic processes in retrieval-induced forgetting*. (Doctoral dissertation, AAI3231330).
- Keresztes, A. & Racsmány, M. (2013). Interference resolution in retrieval-induced forgetting: behavioral evidence for a nonmonotonic relationship between interference and forgetting. *Memory & Cognition*, *41*, 511-518. doi: 10.3758/s13421-012-0276-3
- \* Koessler, S., Engler, H., Riether, C., & Kissler, J. (2009). No retrieval-induced forgetting under stress. *Psychological Science*, *20*, 1356-1363. doi: 10.1111/j.1467-9280.2009.02450.x
- \* Koessler, S., Steidle, L., Engler, H., & Kissler, J. (2012). Stress eliminates retrieval-induced forgetting: Does the oral application of cortisol? *Psychoneuroendocrinology*, *38*, 94-106. doi: 10.1016/j.psyneuen.2012.05.004
- \* Koessler, S., Wöhrmann, C., Zwissler, B., Pfeiffer, A., Ertl, V., & Kissler, J. (2010). Does remembering cause forgetting in chronically stressed people? A study of Ugandan civil war refugees with and without PTSD. *Zeitschrift für Psychologie/Journal of Psychology*, *218*, 71-79. doi: 10.1027/0044-3409/a000014
- \* Koppel, J. (2011). *The contrasting effect of expertise on social contagion and socially shared retrieval-induced forgetting: Explorations across two contexts*. (Doctoral dissertation, AAI3444210).
- Koppel, R. H., & Storm, B. C. (2014). Escaping mental fixation: Incubation and inhibition in creative problem solving. *Memory*, *22*, 340-348. doi: 10.1080/09658211.2013.789914
- \* Koutstaal, W., Schacter, D. L., Johnson, M. K., & Galluccio, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory & Cognition*, *27*, 478-493. doi: 10.3758/BF03211542
- \* Kuhbandner, C., Bäuml, K.-H. T., & Stiedl, F. C. (2009). Retrieval-induced forgetting of negative stimuli: The role of emotional intensity. *Cognition and Emotion*, *23*, 817-830. doi: 10.1080/02699930802204768
- \* Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, *10*, 908-914. doi: 10.1038/nn1918

- \* Kuhl, B. A., Kahn, I., Dudukovic, N. M., & Wagner, A. D. (2008). Overcoming suppression in order to remember: Contributions from anterior cingulate and ventrolateral prefrontal cortex. *Cognitive, Affective & Behavioral Neuroscience*, 8, 211-221. doi: 10.3758/CABN.8.2.211
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.
- \* Lang, J. (2000). *Retrieval failure due to response competition: Inhibition or interference?* (Doctoral dissertation, AAI9943686).
- \* Lechuga, M. T., Moreno, V., Pelegrina, S., Gomez-Ariza, C. J., & Bajo, M. T. (2006). Age differences in memory control: Evidence from updating and retrieval-practice tasks. *Acta Psychologica*, 123, 279-298. doi: 10.1016/j.actpsy.2006.01.006
- Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, 41, 320-331. doi: 10.1016/j.wocn.2013.06.002
- Levy, B. J. & Anderson, M. C. (2002). Repression can (and should) be studied empirically. *Trends in Cognitive Sciences*, 6, 502-503. doi: 10.1016/S1364-6613(02)02025-9
- \* 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, 29-34. doi: 10.1111/j.1467-9280.2007.01844.x
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- \* Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory*, 19, 346-359. doi: 10.1080/09658211.2011.569725
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490. doi: 10.3758/BF03210951
- MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation*. (Vol. 43, pp. 163-214). San Diego, CA: Academic Press
- MacLeod, M. D., & Hulbert, J. C. (2011). Sleep, retrieval inhibition, and the resolving power of human memory. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in Honor of Robert A. Bjork*. (pp. 133-152). New York: Psychology Press
- \* MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, 12, 148-152. doi: 10.1111/1467-9280.00325

- \* 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*, 964-979. doi: 10.1037/0278-7393.31.5.964
- MacLeod, M. D., & Saunders, J. (2008). Retrieval inhibition and memory distortion: Negative consequences of an adaptive process. *Current Directions in Psychological Science*, *17*, 26-30. doi: 10.1111/j.1467-8721.2008.00542.
- \* Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*, 463-473. doi: 10.1037/0022-3514.77.3.463
- \* Macrae, C. N., & Roseveare, T. A. (2002). I was always on my mind: The self and temporary forgetting. *Psychonomic Bulletin & Review*, *9*, 611-614. doi: 10.3758/BF03196320
- \* Malmstrom, T. K. (2004). *Retrieval-induced forgetting in younger and older adults*. (Doctoral dissertation, AAI3130038).
- Marsh, J. E., Sörqvist, P., Beaman, C. P., & Jones, D. M. (2013). Auditory distraction eliminates retrieval induced forgetting. *Experimental Psychology*, *60*, 368-375. 10.1027/1618-3169/a000210.
- Mayberg, H. S. (1994). Frontal-lobe dysfunction in secondary depression. *Journal of Neuropsychiatry and Clinical Neurosciences*, *6*, 428-442.
- \* McCulloch, K. C., Aarts, H., Fujita, K., & Bargh, J. A. (2008). Inhibition in goal systems: A retrieval-induced forgetting account. *Journal of Experimental Social Psychology*, *44*, 857-865. doi: 10.1016/j.jesp.2007.08.004
- McGaugh, J. L. (2000). Neuroscience - Memory - a century of consolidation. *Science*, *287*, 248-251. doi: 10.1126/science.287.5451.248
- McGuire, W. J. (2013). An additional future for psychological science. *Perspectives on Psychological Science*, *8*, 414-423. doi: 10.1177/1745691613491270
- \* Migueles, M., & Garcia-Bajos, E. (2006). Influence of the typicality of the actions in a mugging script on retrieval-induced forgetting. *Psicologica*, *27*, 119-135.
- \* Migueles, M., & Garcia-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*, 1157-1172. doi: 10.1002/acp.1323
- \* Migueles, M., & Garcia-Bajos, E. (2012). The power of script knowledge and selective retrieval in the recall of daily activities. *Journal of General Psychology*, *139*, 100-113. doi: 10.1080/00221309.2012.663817
- Migueles, M. & Garcia-Bajos, E. (2014). The role of exemplar typicality and encoding strategies in category retrieval-induced forgetting. *Memory*, *2*, 212-221. doi: 10.1080/09658211.2013.778288

- Miguez, G., Mash, L. E., Polack, C. W., & Miller, R.R. (2014). Failure to observe renewal following retrieval-induced forgetting. *Behavioural Processes, 103*, 43-51. doi: 10.1016/j.beproc.2013.11.008
- Molenaar, P., C. M., & Campbell, C., G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112-116. doi: 10.1111/j.1467-8721.2009.01619.x
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125. doi: 10.1037/1082-989X.7.1.105
- \* Moulds, M. L., & Kandris, E. (2006). The effect of practice on recall of negative material in dysphoria. *Journal of Affective Disorders, 91*, 269-272. doi: 10.1016/j.jad.2006.01.003
- \* Moulin, C. J. A., Perfect, T. J., Conway, M. A., North, A. S., Jones, R. W., & James, N. (2002). Retrieval-induced forgetting in Alzheimer's disease. *Neuropsychologia, 40*, 862-867. doi: 10.1016/S0028-3932(01)00168-3
- \* Nestor, P. G., Piech, R., Allen, C., Niznikiewicz, M., Shenton, M., & McCarley, R. W. (2005). Retrieval-induced forgetting in schizophrenia. *Schizophrenia Research, 75*, 199-209. doi: 10.1016/j.schres.2005.01.011
- Noortgate, W., & Onghena, P. (2003). Estimating the mean effect size in meta-analysis: Bias, precision, and mean squared error of different weighting methods. *Behavior Research Methods, Instruments, & Computers, 35*, 504-511. doi: 10.3758/bf03195529
- Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review, 114*, 887-953. doi: 10.1037/0033-295X.114.4.887
- \* Odinot, G., Wolters, G., & Lavender, T. (2009). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. *Applied Cognitive Psychology, 23*, 90-97. doi: 10.1002/acp.1443
- \* Ortega, A., Gómez-Ariza, C. J., Román, P., & Bajo, M. T. (2012). Memory inhibition, aging, and the executive deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 178-186. doi: 10.1037/a0024510
- \* Parker, A., & Dagnall, N. (2009). Effects of retrieval practice on conceptual explicit and implicit consumer memory. *Applied Cognitive Psychology, 23*, 188-203. doi: 10.1002/acp.1434
- \* Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1111-1119. doi: 10.1037/0278-7393.28.6.1111

- \* Perfect, T. J., Stark, L.-J., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language*, *51*, 399-417. doi: 10.1016/j.jml.2004.06.003
- \* Potts, R., Law, R., Golding, J. F., & Groome, D. (2012). The reliability of retrieval-induced forgetting. *European Psychologist*, *17*, 1-10. doi: 10.1027/1016-9040/a000040
- \* Quinn, K. A., Hugenberg, K., & Bodenhausen, G. V. (2004). Functional modularity in stereotype representation. *Journal of Experimental Social Psychology*, *40*, 519-527. doi: 10.1016/j.jesp.2003.10.002
- Raaijmakers, J. G. W., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*, 98-122. doi: 10.1016/j.jml.2012.10.002
- \* Raaijmakers, J. G. W., & Jakab, E. k. (2012). Retrieval-induced forgetting without competition: Testing the retrieval specificity assumption of the inhibition theory. *Memory & Cognition*, *40*, 19-27. doi: 10.3758/s13421-011-0131-y
- \* Rafferty, P. *Overloaded and over-stimulated: Impaired retrieval-induced forgetting in borderline personality disorder*. (Doctoral dissertation, AAI3403400).
- Roediger, H. L., III. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*, 261-269.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20-27. doi: 10.1016/j.tics.2010.09.003
- Roediger, H. L., III, & Karpicke, J. D. (2006). The Power of Testing Memory Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*, 181-210. doi: 10.1111/j.1745-6916.2006.00012.x
- \* Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science*, *20*, 1053-1058. doi: 10.1111/j.1467-9280.2009.02415.x
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.
- \* Runnqvist, E., & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition*, *15*, 365-377. doi: 10.1017/S1366728911000034
- \* Rusted, J. M., & Alvares, T. (2008). Nicotine effects on retrieval-induced forgetting are not attributable to changes in arousal. *Psychopharmacology*, *196*, 83-92. doi: 10.1007/s00213-007-0935-4
- \* Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval-induced forgetting and mental imagery. *Memory & Cognition*, *37*, 819-828. doi: 10.3758/MC.37.6.819

- \* Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied*, *8*, 127-142. doi: 10.1037/1076-898X.8.2.127
- \* Saunders, J., & MacLeod, M. D. (2006). Can inhibition resolve retrieval competition through the control of spreading activation? *Memory & Cognition*, *34*, 307-322. doi: 10.3758/BF03193409
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology-Learning Memory and Cognition*, *13*, 501-518. doi: 10.1037//0278-7393.13.3.501
- Schilling, C.J., Storm, B.C., & Anderson, M.C. (2014). Examining the costs and benefits of inhibition in memory retrieval. Unpublished manuscript.
- Schwarzer, G. (2012). meta: Meta-Analysis with R. (Version R package version 2.1-1.).
- \* Sharman, S. J. (2011). Retrieval-induced forgetting of performed and observed bizarre and familiar actions. *Experimental Psychology*, *58*, 361-369. doi: 10.1027/1618-3169/a000103
- \* Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, *2*, 249-253. doi: 10.3758/BF03210965
- Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*. (pp. 175-190): American Psychological Association, Washington, DC.
- Smith, A. D. (1973). Input order and output interference in organized recall. *Journal of Experimental Psychology*, *100*, 147-150. doi: 10.1037/h0035513
- \* Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, *28*, 503-508. doi: 10.3758/BF03201240
- \* Soriano, M. F., Jimenez, J. F., Roman, P., & Bajo, M. T. (2009). Inhibitory processes in memory are impaired in schizophrenia: Evidence from retrieval induced forgetting. *British Journal of Psychology*, *100*, 661-673. doi: 10.1348/000712609X418912
- \* Spitzer, B., & Bäuml, K.-H. T. (2007). Retrieval-induced forgetting in item recognition: Evidence for a reduction in general memory strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 863-875. doi: 10.1037/0278-7393.33.5.863
- \* Spitzer, B., & Bäuml, K.-H. T. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 286-291. doi: 10.1037/a0014363

- \* Spitzer, B., Hanslmayr, S., Opitz, B., Mecklinger, A., & Bäuml, K.-H. T. (2009). Oscillatory correlates of retrieval-induced forgetting in recognition memory. *Journal of Cognitive Neuroscience*, *21*, 976-990. doi: 10.1162/jocn.2009.21072
- \* Starns, J. J., & Hicks, J. L. (2004). Episodic generation can cause semantic forgetting: Retrieval-induced forgetting of false memories. *Memory & Cognition*, *32*, 602-609. doi: 10.3758/BF03195851
- \* Staudigl, T., Hanslmayr, S., & Bäuml, K.-H. T. (2010). Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *Journal of Neuroscience*, *30*, 11356-11362. doi: 10.1523/JNEUROSCI.0637-10.2010
- \* Sterkaj, F. (2012). Retrieval induced forgetting in depression and schizophrenia: Role of mood and the cholinergic system. (Doctoral dissertation).
- \* Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2010). Building consensus about the past: Schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory*, *18*, 170-184. doi: 10.1080/09658210903159003
- Storm, B. C. (2011). The benefit of forgetting in thinking and remembering. *Current Directions in Psychological Science*, *20*, 291-295. doi: 10.1177/0963721411418469
- \* Storm, B. C., & Angello, G. (2010). Overcoming fixation: Creative problem solving and retrieval-induced forgetting. *Psychological Science*, *21*, 1263-1265. doi: 10.1177/0956797610379864
- \* Storm, B. C., Angello, G., & Bjork, E. L. (2011). Thinking can cause forgetting: Memory dynamics in creative problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1287-1293. doi: 10.1037/a0023921
- \* Storm, B. C., Bjork, E., & Bjork, R. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology*, *24*, 617-629. doi: 10.1080/20445911.2012.674030
- \* Storm, B. C., Bjork, E. L., & Bjork, R. A. (2005). Social metacognitive judgments: The role of retrieval-induced forgetting in person memory and impressions. *Journal of Memory and Language*, *52*, 535-550. doi: 10.1016/j.jml.2005.01.008
- \* Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *The Quarterly Journal of Experimental Psychology*, *60*, 909-915. doi: 10.1080/17470210701288706
- \* Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 230-236. doi: 10.1037/0278-7393.34.1.230
- \* Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review*, *13*, 1023-1027. doi: 10.3758/BF03213919

- \* Storm, B. C., & Jobe, T. A. (2012a). Remembering the past and imagining the future: Examining the consequences of mental time travel on memory. *Memory*, *20*, 224-235. doi: 10.1080/09658211.2012.654796
- \* Storm, B. C., & Jobe, T. A. (2012b). Retrieval-induced forgetting predicts failure to recall negative autobiographical memories. *Psychological Science*, *23*, 1356-1363. doi: 10.1177/0956797612443837
- \* Storm, B. C., & Koppel, R. H. (2012). Testing the Cue Dependence of Problem-Solving-Induced Forgetting. *The Journal of Problem Solving*, *4*, 16. doi: 10.7771/1932-6246.1125
- Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, *40*, 827-843. doi: 10.3758/s13421-012-0211-7
- \* Storm, B. C., & Nestojko, J. F. (2010). Successful inhibition, unsuccessful retrieval: Manipulating time and success during retrieval practice. *Memory*, *18*, 99-114. doi: 10.1080/09658210903107853
- Storm, B. C., & Patel, T. N. (2014). Forgetting as a consequence and enabler of creative thinking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- \* Storm, B. C., & White, H. A. (2010). ADHD and retrieval-induced forgetting: Evidence for a deficit in the inhibitory control of memory. *Memory*, *18*, 265-271. doi: 10.1080/09658210903547884
- Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.
- Tempel, T. & Frings, C. (2013). Resolving interference between body movements: retrieval-induced forgetting of motor sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1152-1161. doi: 10.1037/a0030336.
- \* Tomlinson, T. D. (2010). *Learning to forget: An interference theory of cue-independent forgetting*. (Doctoral dissertation, AAI3372994).
- \* Tsukimoto, T., & Kawaguchi, J. (2006). The contribution of category-based and contextual suppression towards retrieval-induced forgetting. *Japanese Psychological Research*, *48*, 40-45. doi: 10.1111/j.1468-5884.2006.00304.x
- Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 145-150. doi: 10.1037/h0023344
- \* Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 315-318. doi: 10.1037/0278-7393.30.2.315
- \* Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition*, *32*, 1265-1272. doi: 10.3758/BF03206317

- Verde, M. F. (2012). Retrieval-induced forgetting and inhibition: A critical review *The psychology of learning and motivation (Vol 56)*. (pp. 47-80): Elsevier Academic Press, San Diego, CA
- \* Verde, M. F., & Perfect, T. J. (2011). Retrieval-induced forgetting in recognition is absent under time pressure. *Psychonomic Bulletin & Review*, *18*, 1166-1171. doi: 10.3758/s13423-011-0143-4
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1-48.
- \* Waldhauser, G. T., Johansson, M., & Hanslmayr, S. (2012). Alpha/beta oscillations indicate inhibition of interfering visual memories. *The Journal of Neuroscience*, *32*, 1953-1961. doi: 10.1523/JNEUROSCI.4201-11.2012
- 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*, 1232-1245. doi: 10.1037/a0030335
- \* Wessel, I., & Hauer, B. J. A. (2006). Retrieval-induced forgetting of autobiographical memory details. *Cognition and Emotion*, *20*, 430-447. doi: 10.1080/02699930500342464
- \* Whitmer, A. J., & Banich, M. T. (2010). Trait rumination and inhibitory deficits in long-term memory. *Cognition and Emotion*, *24*, 168-179. doi: 10.1080/02699930802645762
- \* Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *The American Journal of Psychology*, *114*, 329-354. doi: 10.2307/1423685
- \* Wimber, M., Bäuml, K.-H. T., Bergstrom, Z., Markopoulos, G., Heinze, H.-J., & Richardson-Klavehn, A. (2008). Neural markers of inhibition in human memory retrieval. *Journal of Neuroscience*, *28*, 13419-13427. doi: 10.1523/JNEUROSCI.1916-08.2008
- Wimber, M., Rutschmann, R. M., Greenlee, M. W., & Bäuml, K.-H. T. (2009). Retrieval from episodic memory: Neural mechanisms of interference resolution. *Journal of Cognitive Neuroscience*, *21*, 538-549. doi: 10.1162/jocn.2009.21043
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864
- \* Zellner, M., & Bäuml, K.-H. T. (2005). Intact retrieval inhibition in children's episodic recall. *Memory & Cognition*, *33*, 396-404. doi: 10.3758/BF03193058

### Footnotes

- <sup>1</sup> Some studies omitted the traditional study phase in order to examine memory impairment in semantic memory (e.g., Johnson & Anderson, 2004). We included these studies in our meta-analysis.
- <sup>2</sup> In some studies, only the standard error value, not standard deviation value for each group was available. In these instances, we estimated the standard deviation by multiplying it with the square-root of the number of participants. Although standard errors can be computed in other ways with within-subjects designs (Loftus & Masson, 1994), these alternative procedures yield identical standard error estimates across conditions. In our dataset, no studies that reported standard errors instead of standard deviations presented such identical standard error values across conditions.
- <sup>3</sup> It should be noted that although the high-integration condition did not show a significant retrieval-induced forgetting effect, it did not show a significant retrieval-induced facilitation effect either. In fact, the sign of the averaged effect was opposite from that of facilitation.
- <sup>4</sup> In order to examine the effect of output interference, Ford et al. (2004) intentionally included a condition that examined RIF when Rp+ items were presented first in a recognition memory task. This is the only study that used a recognition memory task that did not control for output interference.
- <sup>5</sup> It should be noted that the association between the extent Rp+ items are strengthened and Rp- items are forgotten can be problematic to examine even within an individual study. One problem, for example, is that the benefits of retrieval practice may be determined in part by the ability to inhibit nontarget competitors. Another problem is that differences between the recallability of items within each category for different participants can lead to a spurious correlation. Specifically, recall performance for Rp- items is likely to be determined not just by inhibition, but by intrinsic or idiosyncratic factors related to those items, including the extent to which a given subject has expertise or familiarity with a given category. Thus, subjects who recall Rp+ items at a higher rate are also more likely to recall Rp- items at a higher rate, completely independent of whether the Rp- items were more susceptible to RIF.

**Table 1**  
*Effect Sizes as a Function of Learning Materials and the Control of Output Interference*

Analysis	$k$	Mean weighted $g$	95% CI	$Q$	$I^2$
All studies	512	0.35**	[0.32, 0.38]	1381.2**	63.0%
Output interference not controlled	256	0.50**	[0.45, 0.55]	602.7**	57.7%
Output interference controlled	280	0.22**	[0.19, 0.26]	670.3**	58.4%
Studies with retrieval practice	472	0.35**	[0.32, 0.38]	1259.5**	62.6%
Output interference not controlled	238	0.50**	[0.45, 0.54]	534.1**	55.6%
Output interference controlled	257	0.23**	[0.19, 0.27]	621.9**	58.8%
Word list: Overall	286	0.36**	[0.32, 0.40]	708.3**	59.8%
Output interference not controlled	144	0.50**	[0.44, 0.56]	309.5**	53.8%
Output interference controlled	158	0.25**	[0.20, 0.9]	366.5**	57.2%
Person trait: Overall	31	0.31**	[0.20, 0.43]	75.0**	60.0%
Output interference not controlled	23	0.33**	[0.20, 0.46]	62.4**	64.7%
Output interference controlled	8	0.26*	[0.08, 0.42]	11.9	41.3%
Visual stimuli: Overall	37	0.28**	[0.15, 0.40]	127.0**	71.6%
Output interference not controlled	14	0.48**	[0.19, 0.78]	44.7**	70.9%
Output interference controlled	23	0.19**	[0.07, 0.32]	73.3**	70.0%
Action: Overall	14	0.47**	[0.25, 0.70]	13.7	5.3%
Output interference not controlled	12	0.53**	[0.28, 0.77]	12.7	13.2%
Output interference controlled	4	0.37	[-0.06, 0.81]	1.9	0.0%
Autobiographical memory: Overall	19	0.40**	[0.26, 0.54]	29.9*	39.7%
Output interference not controlled	14	0.41**	[0.24, 0.57]	21.3	38.9%
Output interference controlled	5	0.39**	[0.13, 0.66]	8.6	53.3%
Factual propositions: Overall	32	0.33**	[0.24, 0.42]	36.1	14.0%
Output interference not controlled	14	0.38**	[0.26, 0.50]	12.6	0.0%
Output interference controlled	24	0.27**	[0.16, 0.38]	27.6	16.6%
Texts: Overall	80	0.32**	[0.21, 0.43]	362.0**	78.2%
Output interference not controlled	35	0.64**	[0.46, 0.82]	141.8**	76.0%
Manipulated high-integration	10	0.14	[-0.08, 0.35]	15.3	41.2%
Manipulated low-integration	10	0.31**	[0.12, 0.49]	11.5	21.7%
Output interference controlled	45	0.08	[-0.04, 0.21]	154.6**	71.5%

Manipulated high-integration	7	0.01	[-0.15, 0.17]	4.6	0.0%
Manipulated low-integration	7	0.25**	[0.09, 0.42]	6.0	0.0%

---

*Note.* \*\*  $p < .01$  \*  $p < .05$ .  $k$  = number of samples.

**Table 2***Effect Sizes for Samples Employing Independent-cuing Paradigms*

Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Independent-cuing paradigm: Overall	67	0.16**	[0.09, 0.24]	164.3**	59.8%
Category-cued independent recall test	24	0.18**	[0.08, 0.28]	40.2*	42.8%
Item-specific independent recall tests	43	0.16**	[0.06, 0.26]	122.1**	65.6%
Cross-category paradigm	36	0.20**	[0.11, 0.29]	73.5**	52.4%
Extralist-cuing paradigm	37	0.18**	[0.06, 0.30]	130.4**	72.4%

*Note.* \*\*  $p < .01$  \*  $p < .05$ . *k* = number of samples.

**Table 3***Effect Sizes in Samples Investigating Retrieval Specificity*

Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Restudy: Overall	17	0.07	[-0.03, 0.17]	8.9	0.0%
Output interference not controlled	3	0.13	[-0.11, 0.37]	0.2	0.0%
Output interference controlled	15	0.05	[-0.05, 0.16]	8.8	0.0%
Category retrieval: Overall	7	0.37**	[0.04, 0.71]	23.0**	73.9%
Output interference not controlled	2	0.46	[-0.75, 1.67]	11.2**	91.1%
Output interference controlled	5	0.34*	[0.01, 0.67]	11.7*	65.9%

*Note.* \*\*  $p < .01$  \*  $p < .05$ . *k* = number of samples.

**Table 4***Effect Sizes as a Function of Type of Retrieval Practice*

Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Provision of retrieval practice feedback: Overall	43	0.29**	[0.19, 0.31]	103.2**	59.5%
Output interference not controlled	18	0.51**	[0.36, 0.66]	23.8	28.5%
Output interference controlled	27	0.18**	[0.07, 0.30]	64.3**	59.6%
Impossible retrieval practice	11	0.28**	[0.15, 0.40]	19.3*	48.2%

*Note.* \*\*  $p < .01$  \*  $p < .05$ . *k* = number of samples. Impossible retrieval paradigm always controls for output interference.

**Table 5**

*Effect Sizes as a Function of Taxonomic Frequency and for Samples that Employed an Episodic Learning Paradigm.*

Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Strong exemplars: Overall	12	0.51**	[0.35, 0.68]	8.7	0.0%
Output interference not controlled	10	0.53**	[0.35, 0.72]	8.2	0.0%
Output interference controlled	4	0.49**	[0.20, 0.78]	1.1	0.0%
Weak exemplars: Overall	12	0.18*	[0.03, 0.34]	10.9	0.0%
Output interference not controlled	10	0.30**	[0.13, 0.47]	7.7	0.0%
Output interference controlled	4	0.00	[-0.37, 0.39]	5.3	43.2%
Episodic learning: Overall	175	0.33**	[0.27, 0.39]	517.9**	66.4%
Output interference not controlled	82	0.47**	[0.38, 0.56]	239.9**	66.2%
Output interference controlled	99	0.21**	[0.14, 0.28]	239.2**	59.0%

*Note.* \*\*  $p < .01$  \*  $p < .05$ . *k* = number of samples.

**Table 6***Effect Sizes for Different Age Groups and Special Populations and their Controls*

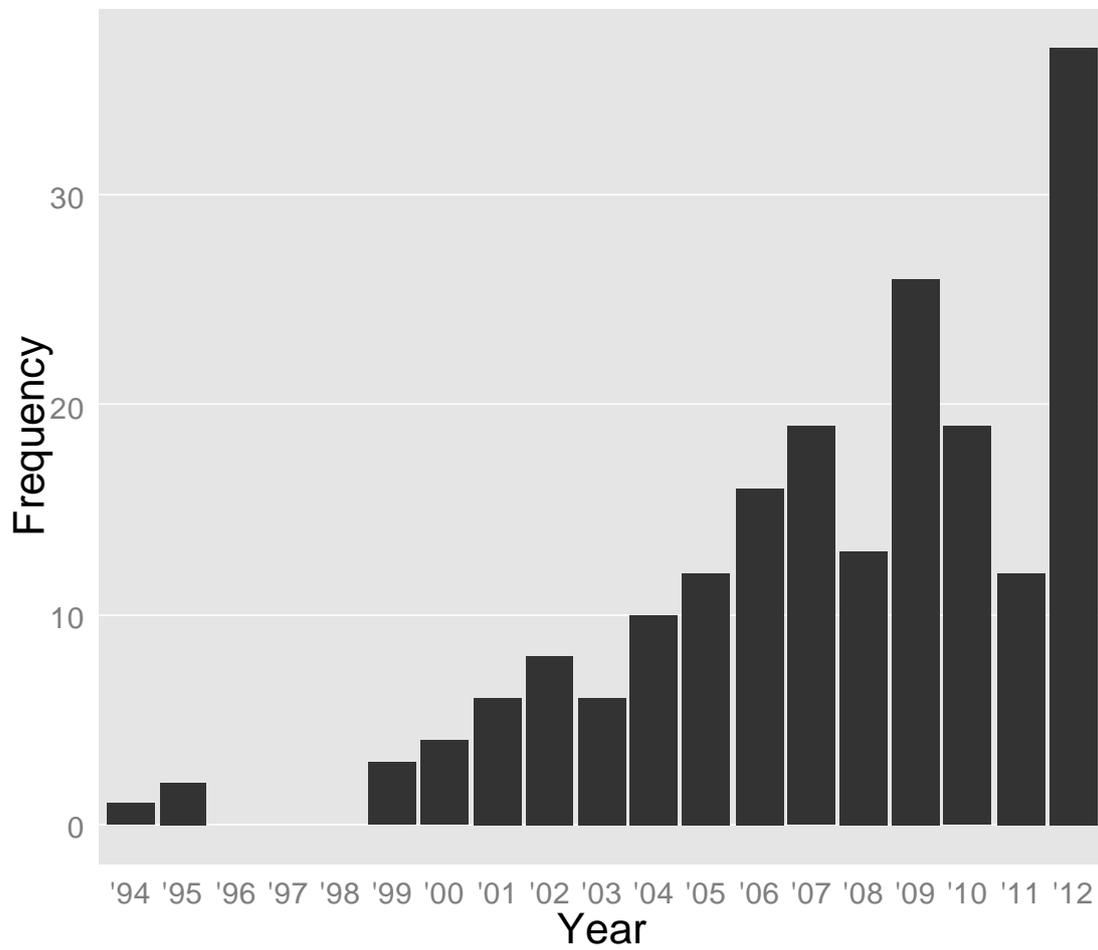
Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Older adults (with a control)	13	0.42**	[0.27, 0.57]	15.7	23.4%
Output interference not controlled	4	0.70**	[0.25, 1.14]	8.5*	64.5%
Output interference controlled	9	0.35**	[0.20, 0.50]	5.0	0.0%
Control younger adults	13	0.47**	[0.30, 0.64]	20.0	39.9%
Output interference not controlled	4	0.64**	[0.20, 1.08]	9.0*	66.5%
Output interference controlled	9	0.42**	[0.24, 0.60]	10.5	23.7%
Young children (with a control)	10	0.56**	[0.37, 0.74]	13.1	31.1%
Output interference not controlled	7	0.61**	[0.43, 0.79]	5.1	0.0%
Output interference controlled	3	0.41	[-0.08, 0.91]	6.0*	66.7%
Control younger adults	6	0.64**	[0.37, 0.92]	8.7	42.7%
Output interference not controlled	4	0.71**	[0.29, 1.14]	8.2*	63.3%
Output interference controlled	2	0.57**	[0.21, 0.93]	0.5	0.0%
Executive function deficits (with a control)	14	0.41**	[0.15, 0.66]	36.2**	64.1%
Output interference not controlled	12	0.55**	[0.26, 0.83]	29.1**	62.2%
Output interference controlled	5	-0.03	[-0.26, 0.20]	5.1	21.1%
Control healthy group	17	0.62**	[0.41, 0.83]	37.6**	57.5%
Output interference not controlled	15	0.69**	[0.47, 0.91]	28.9**	51.5%
Output interference controlled	5	0.22	[-0.10, 0.54]	8.7	54.0%

*Note.* \*\*  $p < .01$  \*  $p < .05$ .  $k$  = number of samples. Impossible retrieval paradigm always controls for output interference.

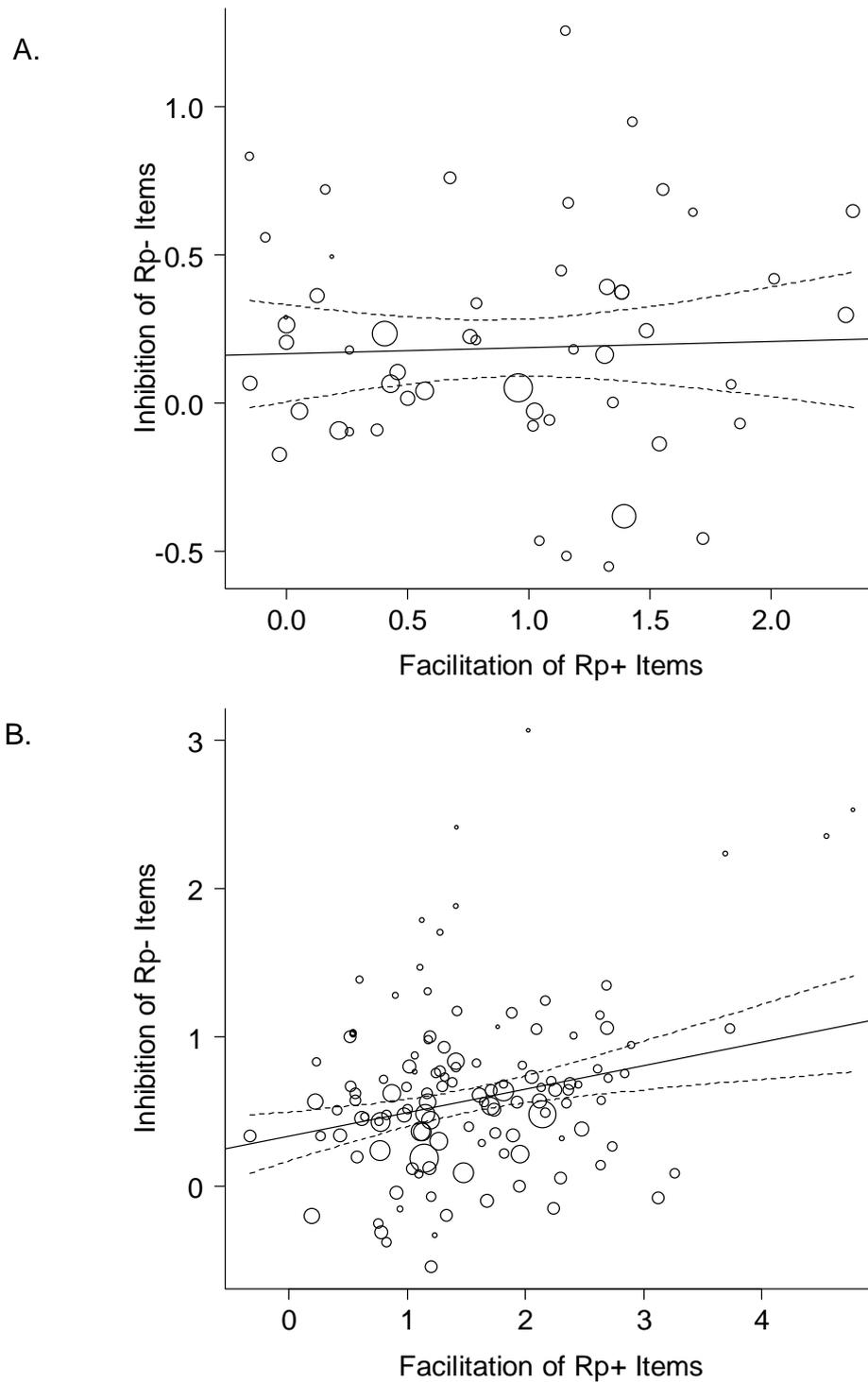
**Table 7***Analysis of Other Moderators*

Analysis	<i>k</i>	Mean weighted <i>g</i>	95% CI	<i>Q</i>	<i>I</i> <sup>2</sup>
Category-cued recall: Overall	230	0.47**	[0.42, 0.51]	488.2**	53.1%
Category-plus-stem cued recall: Overall	112	0.27**	[0.21, 0.33]	243.6**	54.4%
Implicit memory task: Overall	42	0.18**	[0.07, 0.28]	99.3**	58.7%
Output interference not controlled	8	0.55**	[0.23, 0.87]	16.1*	56.4%
Output interference controlled	34	0.12*	[0.02, 0.22]	70.4**	53.1%
Reaction time measure	24	0.18**	[0.07, 0.28]	37.9*	39.4%
Recognition memory performance: Overall	58	0.32**	[0.25, 0.39]	112.7**	49.4%
Extralist paradigm: Overall	28	0.29**	[0.21, 0.37]	55.5**	51.3%
Category cued	6	0.30**	[0.18, 0.41]	5.0	0.0%
Category-stem cued	18	0.28**	[0.17, 0.39]	48.0**	64.6%
Design					
Between-subject design	37	0.51**	[0.31, 0.70]	142.9**	74.8%
Within-subject design	479	0.34**	[0.31, 0.37]	1258.0**	62.0%
Delay: Overall					
Less than 24 hours	419	0.37**	[0.34, 0.40]	974.0**	57.1%
24 hours or more	38	0.09	[-0.07, 0.25]	163.5**	77.4%
Delay: Without text materials					
Less than 24 hours	371	0.35**	[0.32, 0.38]	802.1**	53.9%
24 hours or more	21	0.40**	[0.20, 0.61]	73.7**	72.8%
No retest	15	0.35**	[0.12, 0.59]	48.2**	71.0%
Retest	6	0.55*	[0.08, 1.01]	23.9**	79.1%

*Note.* \*\*  $p < .01$  \*  $p < .05$ . *k* = number of samples.



**Figure 1.** Number of published articles and dissertations on retrieval-induced forgetting available in PsycINFO (searched via the keyword “retrieval-induced forgetting”) as a function of year of publication.



**Figure 2.** Meta-regression analysis predicting the RIF effect from the facilitation effect (standardized difference between Rp+ and Nrp performance) for samples that controlled output interference (A) and those that did not control output interference (B). Circles represent independent samples and the area of circle is proportional to that study's weight ( $w_i$ ).