BRIEF REPORT

Emotional memories are (usually) harder to forget: A meta-analysis of the item-method directed forgetting literature

Kelsi J. Hall[1] · Emily J. Fawcett[2] · Kathleen L. Hourihan[1] · Jonathan M. Fawcett[1]

Abstract

The current meta-analysis explored whether emotional memories are less susceptible to item-method directed forgetting than neutral memories. Basic analyses revealed superior memory for remember (R) than forget (F) items in both the neutral, $M = 19.6\%$, $CI_{95\%}$ [16.1, 23.1], and the emotional, $M = 15.1\%$, $CI_{95\%}$ [12.4, 17.7], conditions. Directed forgetting in either valence condition was larger for (a) words than for other stimuli; (b) recall than recognition tests; (c) studies that used recall prior to recognition testing; (d) shorter lists; and (e) studies that included buffer items. Direct comparison of the magnitude of the directed forgetting effect across neutral and emotional conditions within studies revealed relatively diminished directed forgetting of emotional items compared to neutral items, with an average difference of 4.2%, $CI_{95\%}$ [2.0, 6.4]. However, the nature of this finding varied broadly across studies, meaning that whether – and to what degree – emotional memories are more resilient than neutral memories likely depends on the methodological features of the study in question. Moderator analyses revealed larger differences (a) in studies for which the emotional items were more arousing than the neutral items, and (b) when buffer items were included. Together, these findings suggest that emotional memories are often more resilient to intentional forgetting than neutral memories, although further research is necessary to characterize the circumstances under which these differences emerge.

Keywords Intentional forgetting · Emotion · Memory · Meta-analysis

Introduction

Intentional forgetting plays an important role in cognition by preventing unwanted thoughts from monopolizing our limited-capacity cognitive resources (e.g., Fawcett & Hulbert, 2020). However, recent evidence suggests that some memories are less easily forgotten than others (e.g., Hauswald et al., 2010). Emotional experiences in particular appear to be more resilient than neutral experiences (Bailey & Chapman, 2012; Hauswald et al., 2010), although this finding has been inconsistent (e.g., Yang et al., 2012). The purpose of the current meta-analysis was to synthesize research exploring the intentional forgetting of emotional material within the context of item-method directed forgetting to determine (a) the typical magnitude and interstudy variability of item-method directed forgetting, (b) whether emotional items are less affected by instructions to forget, and, (c) whether the magnitude of any such difference is moderated by specific study characteristics.

Although several paradigms have been used to study intentional forgetting (for a review, see Anderson & Hanslmayr, 2014), we intend to focus exclusively on item-method directed forgetting. In this paradigm, participants are presented with a series of items, each followed by an instruction to Remember (R) or Forget (F). After the presentation of all items, memory performance is tested using recall or recognition. Typically, performance is greater for R than F items; this is known as the directed forgetting effect and has often been explained by the selective encoding of R items at study (MacLeod, 1998). Specifically, participants engage in maintenance rehearsal until a memory instruction is received. Following an R instruction, they elaboratively rehearse the item, whereas following an F instruction they cease rehearsal, eliminating the item from the rehearsal set (Basden et al., 1993; Basden & Basden, 1998; Jing et al., 2019).

The manner in which F items are eliminated from the rehearsal set has become a topic of debate, with some arguing

✉ Kelsi J. Hall
kjh402@mun.ca

1 Department of Psychology, Memorial University of Newfoundland, St. John's, NL, Canada

2 Student Wellness and Counselling Centre, Memorial University of Newfoundland, St. John's, NL, Canada

for the involvement of one or more effortful processes (e.g., Fawcett & Taylor, 2008; Zacks et al., 1996). Fawcett and Taylor (2008) demonstrated that intentional forgetting requires more effort than remembering in the period immediately following the memory instruction, as evidenced by slower secondary probe responses (see also Fawcett et al., 2016; but see Lee, 2018). Later work demonstrated processing resources to be re-allocated away from the representation of the F item in working memory during this same period (e.g., Fawcett & Taylor, 2012; Taylor, 2005; Taylor & Fawcett, 2011). Supporting this idea, neuroimaging studies have revealed frontal brain regions distinct from those involved in incidental forgetting are involved in intentional forgetting (for a review, see Anderson & Hanslmayr, 2014; see also Rizio & Dennis, 2013; Gallant et al., 2018). Notably, others have argued that intentional forgetting is a natural consequence of selectively rehearsing the R items while passively disregarding the F items (Basden & Basden, 1996).

Although directed forgetting (DF) is robust using neutral stimuli such as common nouns (e.g., Bjork, 1970; MacLeod, 1975; MacLeod, 1989; Woodward & Bjork, 1971), objects (e.g., Quinlan et al., 2010), videos (e.g., Fawcett et al., 2013), or visual scenes (e.g., Hauswald & Kissler, 2008), there are a growing number of studies using emotionally valenced materials (e.g., Bailey & Chapman, 2012; Berger et al., 2018; Yang et al., 2016). This line of research provides an important link to real-world applications, as keeping unwanted memories from coming to mind is paramount to our cognitive well-being (e.g., Fawcett & Hulbert, 2020). For this reason, the present study evaluates whether the same process that helps us intentionally forget neutral experiences, such as an outdated phone number, may also be applied to emotional experiences.

Past experiments exploring this question in the context of item-method directed forgetting have proven inconsistent. Some studies have found no difference in the magnitude of DF for emotional and neutral words (e.g., Berger et al., 2018; Gallant & Yang, 2014), whereas others have found a smaller or non-significant effect for emotional words (e.g., Bailey & Chapman, 2012; Yang et al., 2016) or sentences (Lee & Hsu, 2013). Studies using emotionally valenced images have produced similarly mixed results, with some showing no difference in the magnitude of DF for emotional and neutral images (e.g., Quinlan & Taylor, 2014; Taylor et al., 2018; Yang et al., 2012), and others showing a smaller or non-significant effect for emotional images (Hauswald et al., 2010; Nowicka et al., 2011; Zwissler et al., 2011).

Arousal has been proposed as one explanation for these discrepancies (e.g., Gallant & Dyson, 2016; Hauswald et al., 2010). Matching valence conditions for arousal can equate DF for emotional and neutral items (e.g., Gallant et al., 2018; Yang et al., 2012), whereas easing these constraints results in reduced DF for the arousing stimuli (e.g., Hauswald et al.,

2010). This fits with research demonstrating that arousal influences memory to a greater extent than valence (Dolcos et al., 2004; Szőllősi & Racsmány, 2020). However, the impact of arousal on DF has not always been replicated, with some studies observing no effect of emotion even when the emotional stimuli were more arousing than the neutral stimuli (e.g., Taylor et al., 2018).

Although there is no consensus as to whether DF is reduced for emotional items, there are several reasons we would predict such a reduction. Emotional items tend to be processed faster (Kissler & Herbert, 2013), capture attention more easily (Hindi Attar & Müller, 2012), and lead to enhanced memory (Adelman & Estes, 2013). Increased processing of items – preceding (Hourihan & Taylor, 2006) or following (Lee et al., 2007) the memory instruction – reduces item-method DF (e.g., Hauswald et al., 2010; Hourihan & MacLeod, 2008). Consistent with this finding, neuroimaging studies using negative pictures have identified a negative correlation between neural markers of enhanced encoding, and the magnitude of the directed forgetting effect (Hauswald et al., 2010).

In summary, the purpose of the current meta-analysis was to (a) determine whether emotional memories are truly more resistant to intentional forgetting than neutral memories, and, (b) investigate whether the magnitude of any such difference is moderated by study characteristics, including whether the emotional items are more arousing than the neutral items.

Method

Literature search

We conducted a search of the online resources Google Scholar (full-text searched), PsycINFO, PsycARTICLES, and PubMED (title, keywords, and abstract searched) using the following Boolean search phrase: ("item-method" OR "item method") AND ("directed forgetting" OR "intentional forgetting") AND ("emotion" OR "emotional" OR "valence" OR "negative" OR "positive"). The search was conducted until November 2018, restricted to English-language articles, and supplemented by reference review and expert consultation. Authors of included studies were also contacted for raw data and access to missed or unpublished studies.

Study inclusion criteria

Articles reporting at least one estimate of item-method directed forgetting as measured by recall or recognition, within a non-clinical population, using emotional images, words, faces, or events were considered for inclusion. Articles were excluded if they (a) used only clinical samples; (b) reported no experimental data; (c) used a different task (e.g., list-method directed forgetting); (d) reported only samples with a mean

age < 17 years or > 40 years; (e) did not include an emotional condition; (f) were in a language other than English; (g) were inaccessible and the corresponding author failed to provide a copy; (h) reported only an animal model; (i) reported data from an already included source; and/or, (j) provided insufficient information to calculate effect sizes. Studies including multiple comparisons, only some of which were eligible for inclusion (e.g., an article with both young adult and elderly samples) were still included, although only the eligible comparisons were incorporated into our analyses. All studies intermixed the emotional and neutral items, rather than presenting them in blocked or pure lists.

Data extraction

The first author coded each article in consultation with the remaining authors; all coding decisions were documented and discussed until a consensus emerged. In this manner we also coded methodological features for use as moderators, including the stimulus type (words, other), memory task (recall, recognition), average valence and arousal of the stimulus set, whether participants engaged in a recall task prior to the recognition task (if included), list size (the total number of items included in the study phase), whether buffer items were included preceding and/or following the study items (buffers, no buffers), and whether or not a secondary measure (e.g., EEG) was gathered concurrent to the study portion of the task. Because the scale used for valence and arousal varied across studies, these values were standardized by dividing each by the maximum possible value for that study, producing a value ranging from 0 to 1. Moderator analyses applied to differences in the magnitude between neutral and emotional conditions used instead the difference in the arousal and valence ratings for those conditions (reflecting the degree to which emotional items were more arousing or valenced than the neutral items).

Effect size calculation

Effect sizes were calculated as raw mean differences using the equations appropriate for within-subject designs provided by Borenstein et al. (2010, Chapter 4).[1] We first calculated differences estimating the magnitude of DF (R – F) separately for the neutral, negative, and positive conditions within each experiment. A difference was then calculated for the combination of the negative and positive conditions, which we refer to as the emotional condition.[2] We next calculated the difference in the magnitude of DF for each condition. We did this by subtracting performance for the R and F items for each and

then subtracting the resulting values. For comparisons aggregating DF within each valence condition, positive values indicate greater performance for R items compared to F items; for comparisons between the magnitude of DF across valence conditions (e.g., Neutral – Emotional), positive values indicate greater DF in the neutral condition, except for the comparison between the positive and negative conditions (i.e., Positive – Negative) wherein positive values indicate greater DF in the positive condition. Our focus was on the combined (positive + negative) emotional condition, as it made best use of the available data, and because each of the comparisons between the neutral and remaining conditions produced qualitatively similar results.

For recall, the mean percentage of items recalled in each condition was used; for recognition, we instead used the percentage of "hits" for old items. Our preference was to analyze recall accuracy and "hits" because they share a common scale and could therefore be aggregated using raw difference scores. Nonetheless, supplementary analyses were also undertaken to verify whether findings observed in the main analyses were also present using a measure of sensitivity (d').[3] We used values of d' reported in-text or derived from raw data where possible, with the remaining values calculated using aggregate hits and false alarms.

Throughout, standard deviations were required to estimate the standard error of the difference scores. In cases where standard deviations were unavailable from the text or study's authors, we imputed the relevant value as the average of the standard deviations available for that dependent measure. Only a small number of standard deviations were imputed in this manner for our analysis of recall accuracy and recognition "hits"; however, few studies reported d' directly, requiring us to impute most of the standard deviations used in the calculation of those models. For that reason, additional sensitivity analyses were undertaken for our d' models weighting studies by sample size. Sample-size weighted models produced results qualitatively similar to the models using standard errors.

Following calculation of the effect sizes, separate Bayesian random- and mixed-effects models were generated using the *brms 2.9.0* (Bürkner, 2017, 2018) package in *R 3.6.1* (R Core Team, 2018). Each model incorporated random intercepts accounting for variability across samples and dependencies across measured effects. Because several studies reported both recall and recognition, we included a random slope in our analysis of task effects. Models were fit and evaluated for convergence using standard practices (e.g., R-hat < 1.01; Gelman & Hill, 2007). We direct interested readers to past work from our laboratory for more information on our

---

[1] For studies not reporting correlations for our calculations, we estimated these values using the available data. Sensitivity analyses revealed a similar pattern of findings across a range of correlations.
[2] Corresponding valence and arousal ratings were calculated as the average of those conditions. Positive valence ratings were inverted prior to averaging.

---

[3] Because DF for "hits" was calculated as the difference between the R and F items within a given valence condition, which shared a common false alarm rate, we expected that differences between these conditions were already driven in large part by sensitivity.

modelling approach (e.g., Fawcett et al., 2016; Fawcett & Ozubko, 2016).

The priors for our primary analyses are summarized below, but more detailed information is available from the senior author. When comparing the magnitude of directed forgetting within a given valence, our prior expectations relating to the intercept of each model assumed that the average effect in a typical sample should range between -30% and 30%. We further assumed the standard deviation pertaining to random effects should range between 0% and 30%; this broadly permits the "true" effect within any given sample to vary anywhere from -90% to 90%. Our prior for slopes within the moderator models were represented by a normal distribution centred at 0 with a standard deviation of 30.

For the comparison of the magnitude of directed forgetting across valences, our prior expectations relating to the intercept of each model assumed that the average effect in a typical sample should range somewhere between -20% and 20%. We further assumed that the standard deviations pertaining to random effects should range between 0% and 20%; this broadly permits the "true" effect within any given sample to vary anywhere from -60% to 60%. Our prior for slopes within the moderator models were reflected by a normal distribution centred at 0 with a standard deviation of 20.

Due to inconsistent reporting across studies, we fit a separate model for each moderator. Further, heterogeneity was quantified within each of our non-moderator models using prediction intervals (IntHout et al., 2016), which reflect the range of probable "true" effects that would be expected should a new study be conducted like those included in the analysis. For each moderator, Bayesian p-values were calculated reflecting our confidence in the direction of the observed effect (e.g., a value of .95 pertaining to a positive effect would indicate that we are 95% confident that the effect is positive).

Results

Description of studies

Of the 607 studies identified, 31 were included in the final sample (see Table 1), providing 36 effect sizes (see Fig. 1). References contributing to our analyses are marked by an asterisk (*).

Directed forgetting for neutral and emotional conditions

As depicted in Fig. 2, DF was numerically largest in the neutral condition, whereas positive and negative effects were of similar magnitude. These comparisons are explored further below.

Despite clear evidence of an effect in a "typical" study, prediction intervals revealed heterogeneity across studies. Of

these prediction intervals, all but the positive valence condition excluded negative values, indicating that the "true" effect for most studies with methods like those in the present analysis should indicate at least some degree of DF, although some of those effects may be close to 0%.

As summarized in Table 2, the effect of each moderator was consistent across the neutral and emotional conditions. In particular, the magnitude of DF was greatest for comparisons (a) using words than other stimuli; (b) using recall than recognition (this was less convincing for the neutral condition); (c) using recognition for which recognition was preceded by recall; (d) for shorter lists; and (e) for which the study phase included buffers. Minimal evidence was observed supporting the remaining moderators. These effects persisted when measured using d', although the effect of list size was no longer credible for either the neutral or emotional model (confidence dropped to .88 and .83, respectively).

Comparing the magnitude of directed forgetting across conditions

As depicted in Fig. 3, DF tended to be smaller for emotional than neutral items but did not differ between negative and positive items. Prediction intervals again revealed heterogeneity between studies. Of particular interest, prediction intervals for the neutral-emotional comparison revealed the probable "true" effects for studies with methods like those in the present sample would be expected to range from as low as -5.3% to as high as 13.7%. That is to say that whereas we would expect most (75%) studies to support the claim that emotional stimuli are less likely than neutral stimuli to be forgotten intentionally, this may not always be the case: For the remaining studies, similar or even slightly superior DF would be expected for emotional items.

Table 3 provides insight into when we might expect emotional material to be less likely to be forgotten intentionally. Differences between emotional and neutral items were greatest (a) when the emotional items were more arousing than the neutral items, and (b) when the study phase included buffers. As depicted in Fig. 4, the difference in DF for neutral and emotional items was predicted to be numerically equivalent when arousal was matched perfectly, $M = 0.0\%$, $CI_{95\%}$ [-5.5, 5.6].

However, inspection of Fig. 4 also reveals two influential effects (Gallant et al., 2018; Gallant & Yang, 2014) with arousal ratings matched more closely than the remaining data. These points were confirmed as multivariate outliers using the minimum covariance determinant (Fauconnier & Haesbroeck, 2009). Their influence was addressed in two ways. First, the model was refit excluding those studies; this produced a similar, albeit stronger relation (confidence in the effect became .99). Second, a non-linear model was fit to the data using thin-plate regression splines (Wood, 2003), permitting the possibility that performance would asymptote or even invert as arousal was matched between conditions. As also depicted

Table 1 Study characteristics

Article	Stimuli	Task	Recall first	Buffers	Included valences	Arousal ratings			Valence ratings			List size	Secondary measure
						Neu	**Emo**	**Diff**	**Neu**	**Emo**	**Diff**		
Bailey and Chapman (2012, Experiment 1)	Words	Both	Yes	No	Neu, Pos, Neg	–	–	–	–	–	–	240	No
Bailey and Chapman (2012, Experiment 2)	Words	Both	Yes	No	Neu, Pos, Neg	–	–	–	–	–	–	240	Yes
Berger et al. (2018)	Words	Recognition	No	No	Neu, Pos, Neg	0.45	0.64	0.19	0.59	0.21	0.38	90	No
Brandt et al. (2013)	Words	Recognition	No	No	Neu, Neg	–	–	–	–	–	–	80	Yes
Gallant and Dyson (2016)	Words	Recognition	No	No	Neu, Pos, Neg	0.44	0.68	0.24	0.57	0.22	0.35	240	Yes
Gallant and Yang (2014)	Words	Recognition	No	Yes	Neu, Pos, Neg	0.47	0.5	0.03	0.56	0.24	0.32	60	No
Gallant et al. (2018)	Words	Recognition	No	Yes	Neu, Pos, Neg	0.47	0.5	0.03	0.56	0.24	0.32	60	Yes
Gamboa et al. (2017)	Words	Both	Yes	Yes	Neu, Neg	–	–	–	–	–	–	80	No
Hauswald et al. (2010)	Pictures	Recognition	No	No	Neu, Neg	0.31	0.67	0.35	0.58	0.29	0.29	120	Yes
Korfine and Hooley (2000)	Words	Both	Yes	No	Neu, Pos, Neg	0.46	0.60	0.15	0.65	0.24	0.41	42	No
Lee and Hsu (2013, Experiment 1)	Events	Recall	–	No	Neu, Neg	0.48	0.81	.33	0.66	0.31	0.36	60	No
Lee and Hsu (2013, Experiment 2)	Events	Recall	–	No	Neu, Neg	0.48	0.81	.33	0.66	0.31	0.36	60	No
Lee and Hsu (2013, Experiment 3)	Events	Recall	–	No	Neu, Neg	0.48	0.81	.33	0.66	0.31	0.36	60	No
Li et al. (2017)	Events	Recall	–	No	Neu, Pos, Neg	0.43	0.57	0.14	0.57	0.27	0.30	42	No
Liu et al. (2017)	Words	Recognition	No	Yes	Neu, Neg	0.35	0.67	0.32	0.56	0.26	0.30	240	Yes
Marchewka et al. (2016)	Pictures	Recognition	No	No	Neu, Neg	0.17	0.36	0.19	0.62	0.33	0.29	240	Yes
McNally et al. (1998)	Words	Recall	–	Yes	Neu, Pos, Neg	0.35	0.58	0.23	0.61	0.24	0.37	30	No
McNally et al. (1999)	Words	Recall	–	Yes	Neu, Pos, Neg	–	–	–	–	–	–	48	No
Moulds and Bryant (2002)	Words	Recall	–	Yes	Neu, Pos, Neg	–	–	–	0.69	0.16	0.54	30	No
Moulds and Bryant (2008)	Words	Recall	–	Yes	Neu, Pos, Neg	–	–	–	0.69	0.16	0.54	30	No
Nowicka et al. (2011)	Pictures	Recognition	No	No	Neu, Neg	0.42	0.64	0.21	0.67	0.31	0.36	120	Yes
Otani et al. (2012)	Pictures	Recall	–	No	Neu, Pos, Neg	0.29	0.58	0.29	0.52	0.23	0.29	30	No
Patrick et al. (2015)	Words	Recognition	No	No	Neu, Neg	0.46	0.65	0.19	0.50	0.21	0.29	140	Yes
Pierguidi et al. (2016)	Faces	Recognition	No	No	Neu, Pos, Neg	0.36	0.59	0.23	0.56	0.29	0.27	216	Yes
Quinlan and Taylor (2014, Experiment 1)	Faces	Recognition	No	No	Neu, Pos, Neg	–	–	–	–	–	–	60	No
Quinlan and Taylor (2014, Experiment 2)	Faces	Recognition	No	No	Neu, Pos, Neg	–	–	–	–	–	–	60	No
Tay and Yang (2017)	Faces	Recognition	No	No	Pos, Neg	–	0.52	0.02	–	–	–	48	No
Taylor et al. (2018, Experiment 1)	Pictures	Recognition	No	No	Neu, Pos, Neg	0.36	0.55	0.19	0.55	0.27	0.29	96	No
Taylor et al. (2018, Experiment 2)	Pictures	Recognition	No	No	Neu, Pos, Neg	0.36	0.55	0.19	0.55	0.27	0.29	144	No
Tekcan et al. (2008)	Words	Both	Yes	Yes	Neu, Pos, Neg	–	–	–	–	–	–	54	No
Wierzba et al. (2018)	Words	Recognition	No	No	Neu, Neg	0.35	0.62	0.27	0.59	0.41	0.18	120	Yes
Wilhelm et al. (1996)	Words	Both	Yes	Yes	Neu, Pos, Neg	0.37	0.54	0.18	0.62	0.26	0.36	48	No
Yang et al. (2012)	Pictures	Recognition	No	No	Neu, Neg	–	–	–	–	–	–	280	Yes
Yang et al. (2016, Experiment 1)	Words	Recognition	No	Yes	Neu, Neg	–	–	–	–	–	–	64	Yes
Yang et al. (2016, Experiment 2)	Words	Recognition	No	Yes	Neu, Neg	–	–	–	–	–	–	160	Yes
Zwissler et al. (2011)	Pictures	Recognition	No	No	Neu, Pos	0.37	0.53	0.16	0.50	0.20	0.30	36	No

Note. Recall first = whether the recognition test was preceded by a recall test (dashes indicate studies for which only recall was used); Arousal = arousal ratings for neutral and emotion items and the difference between them (dashes indicate studies for which this information was unavailable); Valence = valence ratings for the neutral and emotional items and the difference between them (dashes indicate studies for which this information was unavailable). See text for further details
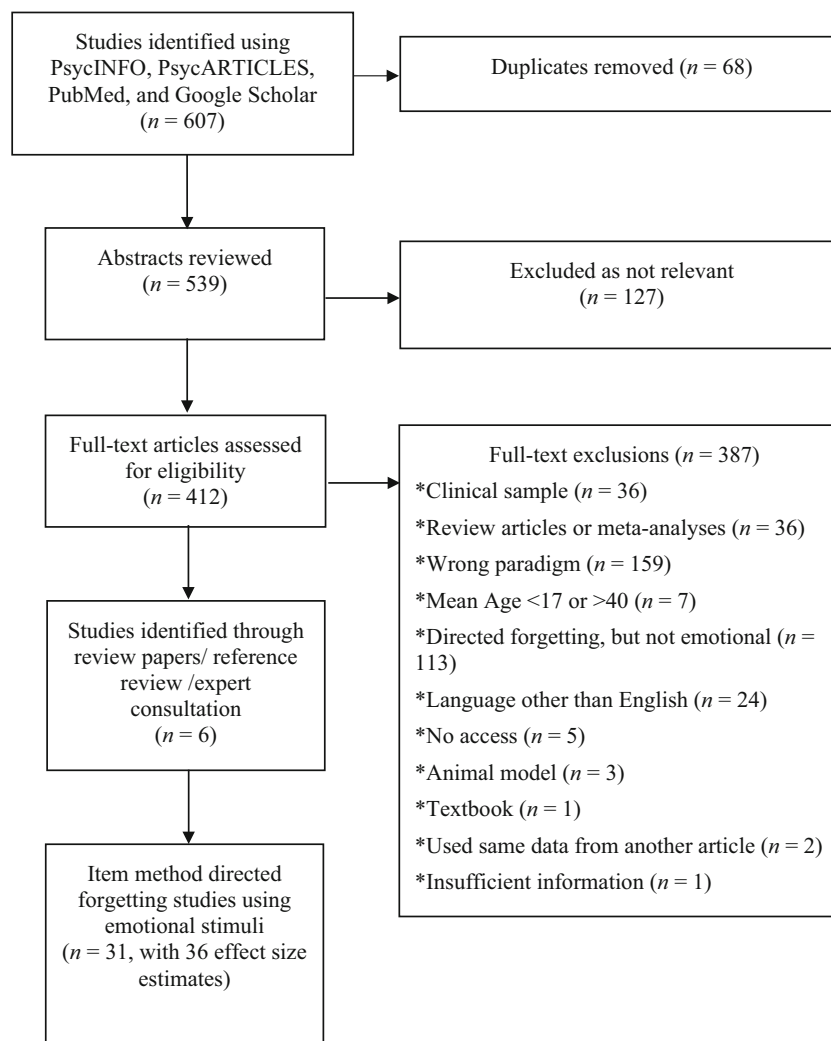
Fig. 1 Meta-analysis inclusion flowchart

in Fig. 4, the non-linear model predicted an initial, rapid decline, leveling off close to 0 (no difference).

Although our analyses of *d'* produced a comparable pattern of differences in the overall magnitude of DF between the emotional and neutral conditions, the same was not true for the moderator analyses. Despite the directionality of the effects matching those described in the preceding paragraph, none of the moderators were credible. This could be caused by the fact that the analysis of *d'* was based on far fewer studies relative to the overall models. Alternately, it is possible that the moderators were driven by differences in response bias. We view this possibility as relatively less likely. Because DF was most often measured as the difference in "hits" for R and F items using a shared false alarm rate, there was little opportunity for response bias to contribute.

Tests of publication bias

Although our analyses provide clear evidence both of a DF effect within each valence condition and a reduction in the magnitude

of DF for emotional (relative to neutral) material, it remains possible that the magnitude of these effects might be driven in part by publication bias, whereby non-supportive findings (particularly with small sample sizes) are preferentially not published. To evaluate this possibility, a series of regression tests were undertaken using the (scaled) standard error or sample size as a moderator, conducted separately for each dependent measure and valence condition (mirroring our moderator analyses, we have limited our models to the neutral and emotional conditions). This approach is analogous to the regression tests included in the *regtest* function of the *metafor* package (Viechtbauer, 2010; see also Sterne & Egger, 2005).

Three of the possible 12 tests demonstrated evidence of a relation between either standard error or sample size and the aggregate effect. Standard error, $B = 7.8$, *CI 95%* [5.4, 10.3], and sample size, $B = -2.3$, *CI 95%* [-4.5, 0.0], predicted the magnitude of DF in the emotional condition, although standard error, $B = 10.1$, *CI 95%* [7.2, 13.3], but not sample size, $B = -1.2$, *CI 95%* [-5.4, 2.1], credibly predicted the magnitude of DF in the neutral condition. Inspection of the funnel plots pertaining to each
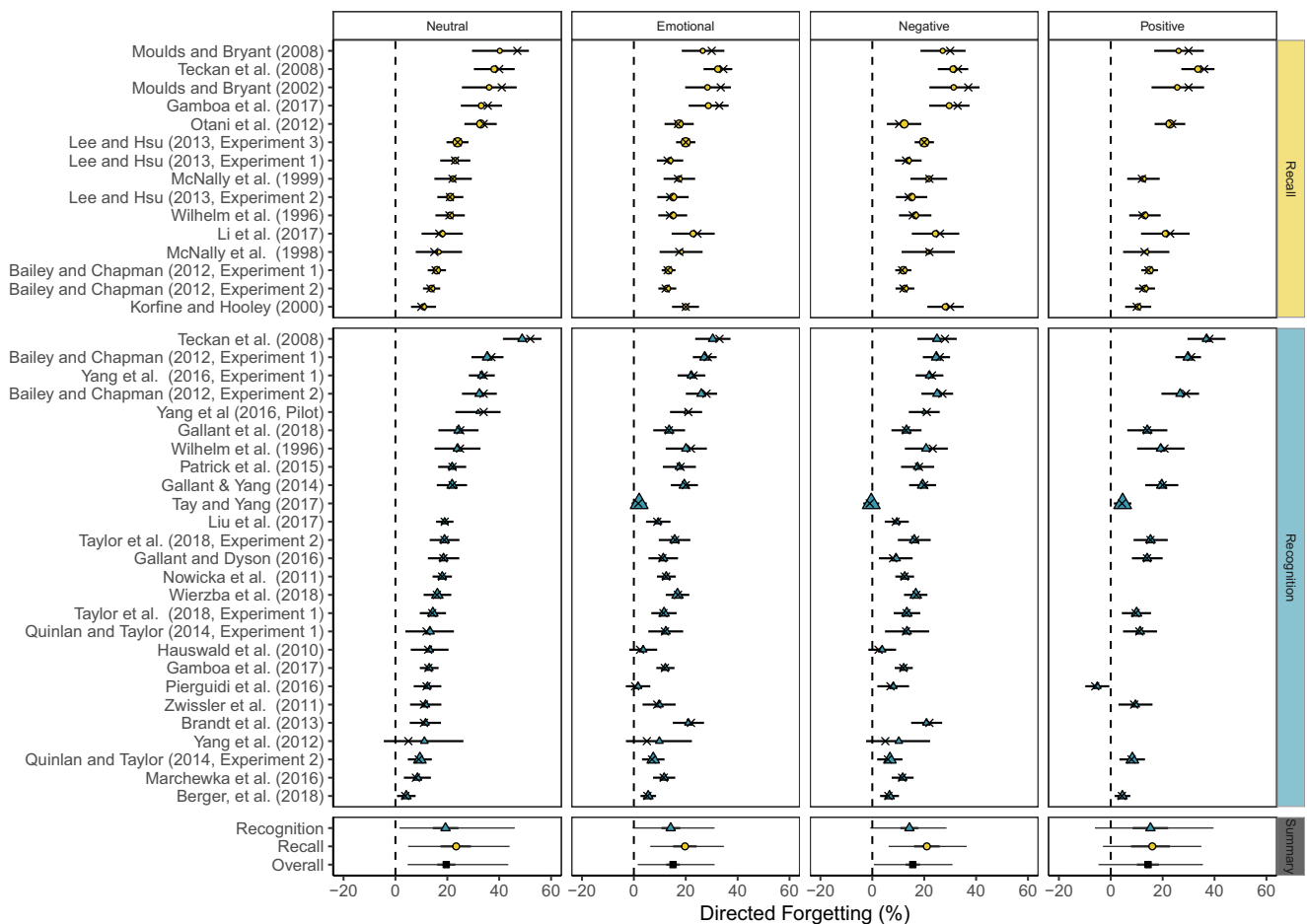
Fig. 2 Mean directed forgetting effect (%) as a function of valence (neutral, emotional, negative, positive) and task type (recall, recognition). *Note.* Yellow Circles: Recall, Blue Triangles: Recognition. Symbols and error bars represent posterior estimates and their corresponding 95% confidence intervals. Xs represent the empirical values reported in the relevant article. Symbol size is scaled to reflect relative sample size. Estimates provided in the bottom panel represent aggregate effects; in this panel, thick lines reflect 95% confidence intervals and thin lines reflect 95% prediction intervals. Data are sorted in descending order according to the effect for the neutral condition

of these analyses demonstrated an apparent tendency for small, imprecise, supportive studies to be published at a rate greater than small, imprecise, non-supportive studies. Importantly, whereas this might suggest some degree of publication bias in our estimate of the magnitude of DF, the neutral and emotional conditions appeared to be influenced to a similar degree. Supporting this assertion, neither standard error, $B = 1.5$, *CI 95%* [-1.1, 4.1], nor sample size, $B = -0.9$, *CI 95%* [-5.1, 3.5], credibly predicted the comparison between emotional and neutral conditions. None of the conducted tests revealed evidence of small sample effects in our analyses of d' (all remaining Bayesian p-values indicated < 75% confidence that the slope tended in a direction suggestive of publication bias). Funnels corresponding to each of the regression tests are provided in the Online Supplementary Material.

Discussion

The current analysis addressed whether emotional memories are more resilient to intentional forgetting than neutral

memories. Initial models demonstrated significant DF for each condition, although these effects were moderated by methodological features. We also demonstrated reduced DF on average for emotional than neutral items, with minimal difference between positive and negative items. Importantly, our data suggest that reduced DF for emotional memories may be driven in part by differences in arousal. Owing to their greater theoretical implications, we begin our discussion with the between-valence comparisons.

Comparing the magnitude of directed forgetting between neutral and emotional items

The present meta-analysis revealed DF to be of lesser magnitude for emotional than neutral items, resolving a previous debate exhibiting equivocal evidence for (Yang et al., 2016) or against (e.g., Taylor et al., 2018) such an effect. In terms of the mechanisms driving this difference, there are several possibilities. First, emotion leads to less efficient performance on tasks requiring cognitive control (e.g., the Stroop or stop-

Table 2 Moderators influencing the magnitude of the directed forgetting effect (%) for neutral and emotional items

Moderator	k	M	Difference (%)	p
Neutral items				
Stimulus				
Words	25	22.3 [17.9, 26.8]	6.1 [-0.5, 13.1]	.96
Other	15	16.1 [10.8, 21.1]		
Task				
Recall	15	23.5 [17.4, 29.1]	-4.0 [-11.5, 3.4]	.86
Recognition	25	19.3 [14.4, 24.3]		
Arousal ratings	25	2.93 [-35.7, 40.7]		.56
Valence ratings	29	19.5 [-23.1, 61.0]		.82
Recall first				
No recall	20	16.2 [11.6, 20.6]	15.0 [5.1, 24.9]	>.99
Recall	5	31.1 [21.9, 40.1]		
List size	40	-3.2 [-6.8, 0.5]		.96
Buffer				
No buffers	25	15.7 [12.2, 19.3]	12.0 [5.7, 18.0]	>.99
Buffers	15	27.6 [22.6, 32.7]		
Secondary measure				
No measure	25	20.7 [16.2, 25.1]	-2.6 [-9.7, 4.4]	.77
Measure	15	18.1 [12.4, 23.6]		
Emotional items				
Stimulus				
Words	25	18.2 [15.1, 21.4]	7.2 [2.5, 11.9]	>.99
Other	16	11.0 [7.5, 14.5]		
Task				
Recall	15	19.7 [15.1, 24.2]	-5.4 [-11.1, 0.1]	.97
Recognition	26	14.2 [10.6, 17.9]		
Arousal ratings	26	0.3 [-27.0, 28.0]		.51
Valence ratings	29	-22.3 [-59.9, 15.7]		.87
Recall first				
No recall	21	11.7 [8.3, 15.0]	12.3 [5.0, 20.0]	>.99
Recall	5	24.0 [17.3, 30.8]		
List size	41	-2.8 [-5.6, 0.0]		.97
Buffer				
No buffers	26	12.6 [9.7, 15.4]	7.7 [2.7, 12.7]	>.99
Buffers	15	20.3 [16.1, 24.5]		
Secondary measure				
No measure	26	16.2 [12.9, 19.6]	-3.0 [-8.5, 2.3]	.87
Measure	15	13.2 [8.9, 17.5]		

Note. k = number of effects; M = mean estimate of directed forgetting for each level of the moderator, in the case of continuous moderators it indicates the slope, 95% confidence interval presented in brackets; Difference (%) = difference in magnitude of directed forgetting between levels of the moderator, 95% confidence interval presented in brackets; p = Bayesian p-value reflecting confidence in the direction of the effect (e.g., p = .95 for a positive effect means 95% confidence the effect is positive)

signal inhibition tasks; Rebetez et al., 2015; Song et al., 2017), so it is possible that active control mechanisms evoked during F trials might have been similarly disadvantaged (e.g.,

Fawcett & Taylor, 2008). Emotional stimuli are also encoded in a manner that is both preferential and more automatic than neutral stimuli (Kissler & Herbert, 2013; Knott et al., 2018; Minor & Herzmann, 2019), potentially converting the maintenance rehearsal thought to occur between item and instruction onset into something more elaborate, frustrating disengagement upon receipt of an F instruction. Related to this point, emotional items also tend to be more semantically interconnected and distinctive than neutral items (Talmi et al., 2007; Talmi & Moscovitch, 2004), either of which would be expected to undermine forgetting (e.g., Golding et al., 1994; Hourihan & Macleod, 2008). Regardless of the mechanism, present findings suggest that not only are emotional memories typically easier to remember, they are also more difficult to forget intentionally.[4]

Despite demonstrating reduced DF relative to neutral items, present findings suggest no difference between positive and negative items. This was unexpected as past research has suggested that negative items are harder to forget intentionally than positive items (Gallant & Dyson, 2016; Otani et al., 2012) and enjoy a stronger enhancement in memory (e.g., Inaba et al., 2005; Szőllősi & Racsmány, 2020). Migita et al. (2011) found such enhancements occur mostly due to pre-attentive processes at item presentation that might undermine DF, as discussed previously. However, Migita et al. (2011) found no enhancement for positive compared to neutral items. The current analysis conflicts with this finding, indicating instead that either valence produces DF of similar magnitude.

Moderators influencing the magnitude of directed forgetting between valence conditions

The present findings offer preliminary support for arousal as one possible explanation for the inconsistent effect of emotion across studies (e.g., Hauswald et al., 2010; Taylor et al., 2018). For example, Bailey and Chapman (2012) compared neutral, negative, and positive items at high and low arousal levels and found reduced DF for high-arousal emotional items in their initial study. Gallant et al. (2018) used arousal-matched items and found no effect of valence on DF. This is not surprising as emotionally arousing information tends to be moved into memory relatively automatically (Kensinger & Corkin, 2004) and has been shown to contribute to greater memory enhancement than emotional valence (Szőllősi & Racsmány, 2020). Emotionally arousing stimuli attract more

---

[4] It is worth mentioning that emotional and neutral items also tend to differ across a variety of other, difficult-to-control dimensions such as frequency or concreteness, which may themselves influence memorability (Fliessbach et al., 2006; Grühn, 2016). However, because few studies report detailed stimulus characteristics, we were unable to evaluate their influence. We strongly encourage future authors to provide their stimuli alongside any publication to ensure analyses such as these are possible in the future. We thank Dr. Ian Neath for this suggestion.
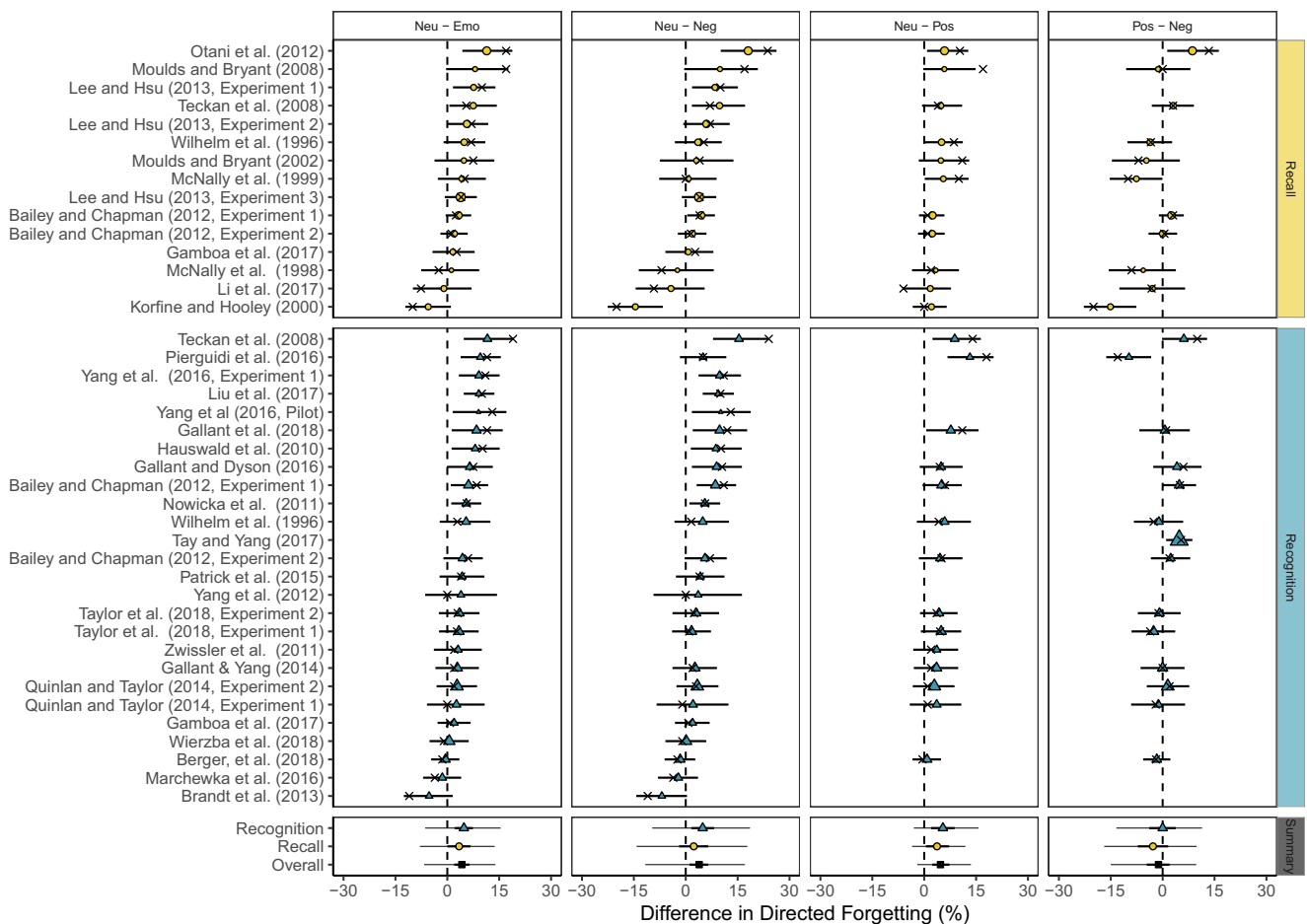
Fig. 3 Mean difference in the magnitude of the directed forgetting effect (%) as a function of valence contrast (neutral-emotional, neutral-negative, neutral-positive, positive-negative) and task type (recall, recognition). *Note.* Yellow Circles: Recall, Blue Triangles: Recognition. Symbols and error bars represent posterior estimates and their corresponding 95% confidence intervals. Xs represent the empirical values reported in the relevant article. Symbol size is scaled to reflect relative sample size. Estimates provided in the bottom panel represent aggregate effects; in this panel, thick lines reflect 95% confidence intervals and thin lines reflect 95% prediction intervals. Data are sorted in descending order according to the effect for the neutral – emotional comparison

attention at encoding than do neutral stimuli (e.g., Simola et al., 2013), and this additional attention may be one of the primary factors in predicting an immediate memory enhancement (Talmi, 2013; Talmi & McGarry, 2012). Given that item-method DF may involve withdrawal of attention from F items (e.g., Fawcett & Taylor, 2010; Taylor, 2005), this withdrawal may therefore be less successful for high-arousal items. Together, these results and the current meta-analysis suggest that reduced DF for emotional items may be partly due to unconstrained differences in arousal rather than differences in valence, although this relationship requires further investigation.

The only other moderator to demonstrate a credible impact on the difference in DF between emotional and neutral conditions was the inclusion of buffer items to mitigate the influence of primacy and recency effects (e.g., Wiswede et al., 2007). However, classic theoretical accounts cannot explain this difference. Primacy effects are historically thought to reflect enhanced rehearsal of

items presented early in the list (Glenberg et al., 1980), but F items should not receive enhanced rehearsal regardless of their serial position. Recency effects are historically attributed to the final items being maintained in working memory until test (Craik, 1970); however, following an F instruction, participants work to push the item from mind (e.g., Fawcett & Taylor, 2008). Other theoretical accounts instead suggest primacy and recency effects occur due to the temporal distinctiveness of the items presented at the beginning and end of the learning phase (Bireta et al., 2018; Neath, 1993a, 1993b), and distinctiveness has been shown to reduce DF (Hourihan & Macleod, 2008). In the context of this comparison, these effects might lead to enhanced memory for both neutral and emotional items presented at the beginning or the end of the study phase. Eliminating these effects would lead to a larger difference between conditions as the emotional items are already processed preferentially relative to the neutral items.

Table 3 Moderators influencing differences in the magnitude of the directed forgetting effect (%) between neutral and emotional conditions

Moderator	K	M	Difference (%)	p
Neu-Emo				
Stimulus				
Words	25	3.5 [0.7, 6.3]	-1.8 [-6.4, 2.6]	.80
Other	15	5.3 [1.8, 8.8]		
Task				
Recall	15	3.4 [0.0, 6.7]	1.4 [-2.6, 5.2]	.76
Recognition	25	4.8 [2.1, 7.4]		
Arousal ratings	25	-21.0 [-44.0, 3.6]		.96
Valence ratings	29	2.3 [-25.4, 29.8]		.56
Recall first				
No recall	20	3.9 [1.1, 6.8]	3.0 [-3.6, 9.7]	.83
Recall	5	7.0 [1.1, 13.0]		
List size	40	0.2 [-2.2, 2.6]		.58
Buffers				
No buffers	25	3.0 [0.5, 5.5]	4.5 [-0.0, 8.9]	.97
Buffers	15	7.5 [3.8, 11.1]		
Activity measure	40			
No measure	25	3.6 [0.7, 6.5]	1.4 [-3.1, 5.8]	.73
Measure	15	5.0 [1.6, 8.3]		
Neg-Pos				
Stimulus				
Words	17	-2.0 [-6.3, 2.1]	-2.2 [-9.5, 4.9]	.74
Other	8	0.2 [-5.5, 6.0]		
Task				
Recall	11	-2.8 [-7.2, 1.6]	2.7 [-2.0, 7.5]	.88
Recognition	14	0.0 [-3.9, 3.8]		
Arousal ratings	14	-11.0 [-49.4, 27.1]		.71
Valence ratings	19	-10.5 [-43.4, 23.1]		.73
Recall first				
No recall	10	-0.5 [-4.1, 3.0]	4.5 [-2.6, 11.5]	.90
Recall	4	3.9 [-2.2, 10.1]		
List size	25	0.7 [-3.1, 4.3]		.65
Buffers				
No buffers	15	-0.7 [-5.1, 3.7]	-1.5 [-8.9, 5.8]	.66
Buffers	10	-2.1 [-8.0, 3.7]		
Activity measure				
No measure	20	-1.6 [-5.6, 2.3]	1.4 [-6.18, 9.08]	.65
Measure	5	-0.2 [-6.8, 6.3]		

Note. k = number of effects; M = mean estimate of directed forgetting for each level of the moderator, in the case of continuous moderators it indicates the slope, 95% confidence interval presented in brackets; Difference (%) = difference in magnitude of directed forgetting between levels of the moderator, 95% confidence interval presented in brackets; p = Bayesian *p*-value reflecting confidence in the direction of the effect (e.g., p = .95 for a positive effect means 95% confidence the effect is positive), *Neu* = Neutral, *Emo* = Emotional, *Neg* = Negative, *Pos* = Positive

Directed forgetting for neutral and emotional conditions

In addition to differences in DF across valences, each valence condition also demonstrated better memory for R compared to F items. This is consistent with previous research observing DF for emotional items (Gallant et al., 2018; Quinlan & Taylor, 2014; Taylor et al., 2018). Although past work would have predicted greater DF for recall than recognition (e.g., Titz & Verhaeghen, 2010), the current analysis demonstrated weak support for this claim in the neutral condition. Nonetheless, there does appear to be a tendency for larger DF as measured using recall. The contrast between words and more elaborative stimuli (e.g., Quinlan et al., 2010; Titz & Verhaeghen, 2010) was more convincing, and could be attributed to the fact that images are often processed in a distinctive manner (Ensor et al., 2019), granting them an advantage relative to words (Nelson et al., 1976). This result supports earlier claims that distinctive encoding attenuates DF (Hourihan & Macleod, 2008).

The present study also found greater DF for shorter lists. This might speculatively be attributed to fatigue known to occur during long tasks requiring repeated bouts of cognitive control (i.e., Stroop; Rauch & Schmitt, 2009). In the context of DF, this could mean that as the task progresses participants rehearse the R items less effectively while also becoming less efficient at pushing the F items from mind. This could lead to both a reduction in memory for R items and a (paradoxical) improvement in memory for F items, reducing the overall magnitude of DF. Further research is needed to evaluate this possibility.

Studies including buffers demonstrated greater DF than those that did not. As discussed previously, items presented at the beginning and end of the study list may be temporally distinct, leading to less DF for those items. Using buffers would work to eliminate this enhancement for F items presented at the beginning and end of the list, producing greater DF as a result.

Among studies using recognition, those that also had a recall test preceding it demonstrated greater DF than those that did not. This could be attributed to testing effects (Roediger & Butler, 2011), as R items have been shown to be recalled prior to F items (e.g., Lee, 2013). This could lead to enhanced memory for R items and therefore greater DF on the recognition test. The recall of the R items might also operate like the retrieval practice phase of a retrieval-induced forgetting paradigm (Anderson et al., 1994), with F items competing for attentional focus, resulting in their down-regulation and failure to be recognized during the following recognition task.

Conclusion

The present meta-analysis demonstrates that emotional memories are typically harder to forget intentionally than neutral
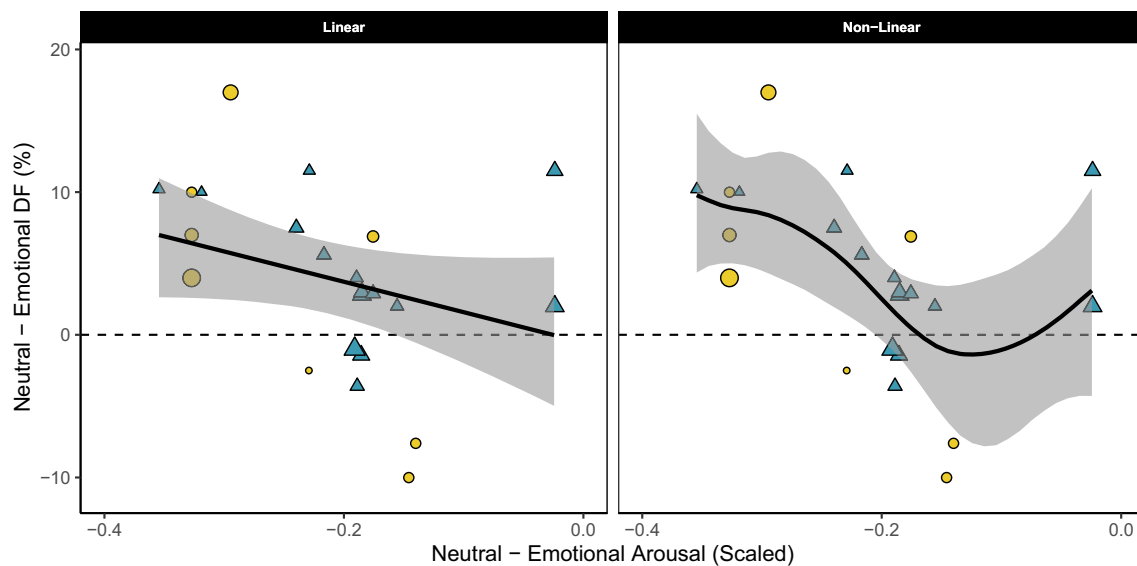
Fig. 4 Difference in the magnitude of the directed forgetting effect (%) between the neutral and emotional conditions as a function of the difference in arousal (neutral – negative) and task type (recall, recognition). *Note.* Yellow Circles: Recall, Blue Triangles: Recognition. X-axis indicates the mean difference in (scaled) arousal between conditions. The Y-axis indicates the difference in the magnitude of directed forgetting between conditions (%). The left panel represents the effect fitted as a linear function; the right panel represents the effect fitted as a non-linear function using thin-plate regression splines. Symbol size is scaled to reflect relative sample size

memories, at least in the context of an item-method directed forgetting paradigm. However, the magnitude of this difference varies broadly from one study to the next, driven by factors such as variation in arousal. Future research should further isolate the role played by these – and other – moderators in determining the impact of emotion on our ability to exert control over unwanted experiences.

References

(References included in our analyses are indicated by an asterisk "*")

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535. https://doi.org/10.1016/j.cognition.2013.08.014

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(5), 1063-1087.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279-292.

*Bailey, K., & Chapman, P. (2012). When can we choose to forget? An ERP study into item-method directed forgetting of emotional words. *Brain and Cognition*, *78*(2), 133–147.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. McLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 603–616.

Basden B. H, Basden D. R. Directed forgetting: further comparisons of the item and list methods. *Memory* 1996;4(6):633-653. https://doi.org/10.1080/741941000

*Berger, N., Crossman, M., & Brandt, K. (2018). No evidence for age-related differences in item-method directed forgetting of emotional words. *Quarterly Journal of Experimental Psychology*, *71*(3), 595-604.

Bireta, T. J., Gabel, A. J., Lamkin, R. M., Neath, I., & Surprenant, A. M. (2018). Distinctiveness and serial position functions in implicit memory. *Journal of Cognitive Psychology*, *30*(2), 222-229.

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

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97-111.

*Brandt, K., Nielsen, M., & Holmes, A. (2013). Forgetting emotional and neutral words: An ERP study. *Brain Research*, *1501*, 21–31.

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

Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395-411. https://doi.org/10.32614/RJ-2018-017

Craik, F. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*(2), 143-148.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *Neuroimage*, 23(1), 64-74.

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory & Cognition*, 47(1), 182-193.

Fauconnier, C., & Haesbroeck, G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology, 6*, 363-379. https://doi.org/10.1016/j.stamet.2008.12.005

Fawcett, J., Lawrence, M., & Taylor, T. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology: General, 145*, 56–81.

Fawcett, J., & Hulbert, J. (2020). The many faces of forgetting: toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory and Cognition, 9*(1), 1-18. https://doi.org/10.1016/j.jarmac.2019.11.002

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology*, 70(2), 99-115. https://doi.org/10.1037/cep0000089

Fawcett, J., & Taylor, T. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 6*, 1168–1181.

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition, 38*(6), 797–808.

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica*, 139(1), 84-90.

Fawcett, J., Taylor, T., & Nadel, L. (2013). Intentional forgetting diminishes memory for continuous events. *Memory, 21*, 675– 694.

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage, 32*(3), 1413-1421. https://doi.org/10.1016/j.neuroimage.2006.06.00

*Gallant, S., & Dyson, B. (2016). Neural modulation of directed forgetting by valence and arousal: An event-related potential study. *Brain Research*, 1648, 306-316.

*Gallant, S. N., Pun, C., & Yang, L. (2018). Age differences in the neural correlates underlying control of emotional memory: An event-related potential study. *Brain Research*, 1697, 83-92.

Gallant, S. N., & Yang, L. (2014). Positivity effect in source attributions of arousal-matched Emotional and non-emotional words during item-based directed forgetting. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.01334

*Gamboa, O. L., Garcia-Campayo, J., Müller, T., & von Wegner, F. (2017). Suppress to forget: The effect of a mindfulness-based strategy during an emotional item-directed forgetting paradigm. *Frontiers in psychology, 8*, 432.

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

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. H., & Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory, 6*(4), 355–369. https://doi.org/10.1037/0278-7393.6.4.355

Golding, J.M, Long, D.L., & MacLeod, C.M. (1994). You can't always forget what you want: Directed forgetting of related words. *Journal of Memory and Language, 33*, 492-510. https://doi.org/10.1006/jmla.1994.1023

Grühn, D. (2016). *An English Word Database of EMOtional TErms (EMOTE). Psychological Reports, 119*(1), 290–308. https://doi.org/10.1177/0033294116658474

Hauswald, A., & Kissler, J. (2008). Directed forgetting of complex pictures in an item method paradigm. *Memory*, 16, 797–809.

*Hauswald, A., Schulz, H., Iordanov, T., & Kissler, J. (2010). ERP dynamics underlying successful directed forgetting of neutral but not negative pictures. *Social Cognitive and Affective Neuroscience*, 6, 450–459.

Hindi Attar, C., & Müller, M. (2012). Selective attention to task-irrelevant emotional distractors is unaffected by the perceptual load associated with a foreground task (selective attention to emotional distractors). *PLoS ONE, 7*(5), e37186.

Hourihan, K., & Macleod, C. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, 62(4), 242-246.

Hourihan, K., & Taylor, T. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1354-1365. https://doi.org/10.1037/0096-1523.32.6.1354

Inaba, M., Nomura, M., & Ohira, H. (2005). Neural evidence of effects of emotional valence on word recognition. *International Journal of Psychophysiology*, 57(3), 165-173.

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7), e010247.

Jing, J., Qi, M., Gao, H., & Zhang, Q. (2019). The role of forgetting cues in directed forgetting: Ceasing maintenance rehearsal. *Acta Psychologica, 199*, 102922.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences*, 101(9), 3310-3315. https://doi.org/10.1073/pnas.0306408101

Kissler, J., & Herbert, C. (2013). Emotion, Etmnooi, or Emitoon?-Faster lexical access to emotional than to neutral words during reading. Biological Psychology, *92*(3), 464-479. https://doi.org/10.1016/j.biopsycho.2012.09.004

Knott, L. M., Howe, M. L., Toffalini, E., Shah, D., & Humphreys, L. (2018). The role of attention in immediate emotional false memory enhancement. *Emotion*, 18(8), 1063-1077. https://doi.org/10.1037/emo0000407

*Korfine, L., & Hooley, J. M. (2000). Directed forgetting of emotional stimuli in borderline personality disorder. *Journal of Abnormal Psychology*, 109(2), 214- 221.

Lee, Y. (2013). Costs and benefits in item-method directed forgetting: Differential effects of encoding and retrieval. *The Journal of General Psychology*, 140(3), 159-173.

Lee, Y. (2018). Withdrawal of spatial overt attention following intentional forgetting: Evidence from eye movements. *Memory, 26*(4), 503-513.

*Lee, Y., & Hsu, Y. (2013). How do we forget negative events? The role of attentional, cognitive, and metacognitive control. *Cognition and Emotion*, 27(3), 401-415.

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

Li, G., Wang, L., & Han Y. (2017). Directed forgetting of negative performed actions is difficult: A behavioural study. *Quarterly Journal of Experimental Psychology, 70*(1), 53–61.

*Liu, T., Chen, N., & Cheng, S. (2017). Selective rehearsal is affected by the emotionality of the encoding context in item-method directed forgetting: An event-related potential study. *Biological Psychology*, 123, 15-24.

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(3), 271- 279.

MacLeod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 13- 21.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches (pp. 144-163)*. Lawrence Erlbaum Associates.

*Marchewka, A., Wypych, M., Michałowski, J., Sińczuk, M., Wordecha, M., Jednoróg, K., & Nowicka, A. (2016). What is the effect of basic emotions on directed forgetting? Investigating the role of basic emotions in memory. *Frontiers in Human Neuroscience*, *10*, 378.

*McNally, R. J., Otto, M. W., Yap, L., Pollack, M. H., & Hornig, C. D. (1999). Is panic disorder linked to cognitive avoidance of threatening information?. *Journal of Anxiety Disorders*, *13*(4), 335-348.

*McNally, R. J., Metzger, L. J., Lasko, N. B., Clancy, S. A., & Pitman, R. K. (1998). Directed forgetting of trauma cues in adult survivors of childhood sexual abuse with and without posttraumatic stress disorder. *Journal of Abnormal Psychology*, *107*(4), 596- 601.

Migita, M., Otani, H., Libkuman, T. M., & Sheffert, S. M. (2011). Preattentive processing, poststimulus elaboration, and memory for emotionally arousing stimuli. *The Journal of General Psychology*, *138*(4), 260-280.

Minor, G., & Herzmann, G. (2019). Effects of negative emotion on neural correlates of item and source memory during encoding and retrieval. *Brain Research*, *1718*, 32-45.

*Moulds, M. L., & Bryant, R. A. (2008). Avoidant encoding in acute stress disorder: a prospective study. *Depression and Anxiety*, *25*(12), E195-E198.

*Moulds, M. L., & Bryant, R. A. (2002). Directed forgetting in acute stress disorder. *Journal of Abnormal Psychology*, *111*(1), 175- 179.

Neath, I. (1993a). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, *32*, 820–840. https://doi.org/10.1006/jmla.1993.1041

Neath, I. (1993b). Distinctiveness and serial position effects in recognition. Memory & *Cognition*, 21, 689–698. https://doi.org/10.3758/BF03197199

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 523- 528.

*Nowicka, A., Marchewka, A., Jednorog, K, Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex*, *21*, 539–549.

*Otani, H., Libkuman, T., Goernert, P., Kato, K., Migita, M., Freehafer, S., & Landow, M. (2012). Emotion, directed forgetting, and source memory. *British Journal of Psychology*, *103*(3), 343-358.

*Patrick, R. E., Kiang, M., & Christensen, B. K. (2015). Neurophysiological correlates of emotional directed-forgetting in persons with Schizophrenia: An event-related brain potential study. *International Journal of Psychophysiology*, *98*(3), 612-623.

*Pierguidi, L., Righi, S., Gronchi, G., Marzi, T., Caharel, S., Giovannelli, F., & Viggiano, M. (2016). Emotional contexts modulate intentional memory suppression of neutral faces: Insights from ERPs. *International Journal of Psychophysiology*, *106*, 1-13.

*Quinlan, C. K., & Taylor, T. L. (2014). "I never forget a face, but in your case I'll be glad to make an exception": Intentional forgetting of emotional faces. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale*, *68*(3), 212–221. https://doi.org/10.1037/cep0000024

Quinlan, C., Taylor, T., & Fawcett, J. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *64*(1), 41-46.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rauch, W., & Schmitt, K. (2009). Fatigue of cognitive control in the Stroop-task. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 31, 750-755.

Rebetez, M., Rochat, L., Billieux, J., Gay, P., & Van Der Linden, M. (2015). Do emotional stimuli interfere with two distinct components of inhibition? *Cognition and Emotion*, *29*(3), 559-567.

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

Roediger, H., & Butler, A. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27.

Simola, J., Torniainen, J., Moisala, M., Kivikangas, M., & Krause, C.M. (2013). Eye movement related brain responses to emotional scenes during free viewing. *Frontiers in Systems Neuroscience, 7*, 1-16. https://doi.org/10.3389/fnsys.2013.00041

Song, S., Zilverstand, A., Song, H., Uquillas, F. D. O., Wang, Y., Xie, C., Chen,. Li., & Zou, Z. (2017). The influence of emotional interference on cognitive control: A meta-analysis of neuroimaging studies using the emotional Stroop task. *Scientific Reports*, *7*(1), 1-9.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 99–110). Wiley.

Szőllősi, Á., & Racsmány, M. (2020). Enhanced mnemonic discrimination for emotional memories: the role of arousal in interference resolution. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-020-01035-3

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science*, *22*, 430-436.

Talmi, D., Luk, B. T., McGarry, L. M., & Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *Journal of Memory and Language*, *56*(4), 555–574.

Talmi, D., & McGarry, L.M. (2012). Accounting for immediate emotional memory enhancement. *Journal of Memory and Language, 66*, 93-108. https://doi.org/10.1016/j.jml.2011.07.009

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition, 32*, 742-751. https://doi.org/10.3758/BF03195864

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *Quarterly Journal of Experimental Psychology*, *58A*(4), 613–629 (Erratum in 58A, 1343).

Taylor, T. L., & Fawcett, J. M. (2011). Larger IOR effects following forget than following remember instructions depend on exogenous attentional withdrawal and target localization. *Attention, Perception, & Psychophysics*, *73*(6), 1790–1814.

*Taylor, T., Quinlan, L., & Vullings, C. (2018). Decomposing item-method directed forgetting of emotional pictures: Equivalent costs and no benefits. *Memory & Cognition*, *46*, 132-147.

*Tay, P. K., & Yang, H. (2017). Angry faces are more resistant to forgetting than are happy faces: directed forgetting effects on the identity of emotional faces. *Journal of Cognitive Psychology*, *29*(7), 855-865.

*Tekcan, A. I., Taş, A. Ç., Topçuoğlu, V., & Yücel, B. (2008). Memory bias in anorexia nervosa: Evidence from directed forgetting. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 369-380.

Titz, C., & Verhaeghen, P. (2010). Aging and directed forgetting in episodic memory: A meta-analysis. *Psychology and Aging*, *25*(2), 405-411.

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

*Wierzba, M., Riegel, M., Wypych, M., Jednoróg, K., Grabowska, A., & Marchewka, A. (2018). Cognitive control over memory–individual differences in memory performance for emotional and neutral material. *Scientific Reports*, *8*(1), 1-18.

*Wilhelm, S., Mcnally, R. J., Baer, L., & Florin, I. (1996). Directed forgetting in obsessive-compulsive disorder. *Behaviour Research and Therapy, 34*, 633-641.

Wiswede, D., Rüsseler, J., & Münte, T. (2007). Serial position effects in free memory recall—An ERP-study. *Biological Psychology*, *75*(2), 185-193.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*, 95-114. https://doi.org/10.1111/1467-9868.00374

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, 89(1), 109.

*Yang, W., Chen, Q., Liu, P., Cheng, H., Cui, Q., Wei, D., Zhang, Q., & Qiu, J. (2016). Abnormal brain activation during directed forgetting of negative memory in depressed patients. *Journal of Affective Disorders, 190*, 880-888.

*Yang, W., Liu, P., Xiao, X., Li, X., Zeng, C., Qiu, J., & Zhang, Q. (2012). Different neural substrates underlying directed forgetting for negative and neutral images: An event-related potential study. *Brain Research*, *1441,* 53-63.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22,* 143–156.

*Zwissler, B., Koessler, S., Engler, H., Schedlowski, M., & Kissler, J. (2011). Acute psycho-social stress does not disrupt item-method directed forgetting, emotional stimulus content does. *Neurobiology of Learning and Memory*, *95*(3), 346-354.