Specificity of Future Thinking in Depression: A Meta-Analysis

Beau Gamble[1] iD, David Moreau[1,2], Lynette J. Tippett[1,2,3], and Donna Rose Addis[1,2,3,4]

[1]School of Psychology, The University of Auckland; [2]Centre for Brain Research, The University of Auckland; [3]Brain Research New Zealand, Auckland, New Zealand; and [4]Rotman Research Institute, Baycrest Health Sciences, Toronto, Canada

Abstract

Reduced specificity of autobiographical memory has been well established in depression, but whether this *overgenerality* extends to future thinking has not been the focus of a meta-analysis. Following a preregistered protocol, we searched six electronic databases, Google Scholar, and personal libraries and contacted authors in the field for studies matching search terms related to depression, future thinking, and specificity. We reduced an initial 7,332 results to 46 included studies, with 89 effect sizes and 4,813 total participants. Random-effects meta-analytic modeling revealed a small but robust correlation between reduced future specificity and higher levels of depression ($r = -.13$, $p < .001$). Of the 11 moderator variables examined, the most striking effects were related to the emotional valence of future thinking ($p < .001$) and the sex of participants ($p = .025$). Namely, depression was linked to reduced specificity for positive (but not negative or neutral) future thinking, and the relationship was stronger in samples with a higher proportion of males. This meta-analysis contributes to our understanding of how prospection is altered in depression and dysphoria and, by revealing areas where current evidence is inconclusive, highlights key avenues for future research.

Keywords

Depression has long been associated with vagueness in recalling the past—often referred to as *overgeneral* autobiographical memory (Williams & Scott, 1988). When prompted to recall a personal event, people with depression tend to retrieve memories that are broad or categorical (e.g., "all the times I've lost at sports") rather than specific to a place and time (e.g., "playing squash at the club last Thursday"). Overgeneral memory in depression is now well established, with moderate-to-large effect sizes reported in two meta-analyses (van Vreeswijk & de Wilde, 2004; Williams et al., 2007). Growing research since the mid-1990s suggests that this overgenerality may extend to various forms of autobiographical future thinking, in line with evidence that memory and future simulation rely on the same brain network (Benoit & Schacter, 2015). For instance, reduced specificity of future thought has been found in suicidal (Williams et al., 1996), currently depressed or remitted (Addis, Hach, & Tippett, 2016), and dysphoric individuals (Dickson & Bates, 2006; MacLeod &

Cropley, 1995). Impairments in depression are evident also on measures that are closely associated with the specificity of future thinking, such as the level of episodic detail in future simulations (King, MacDougall, Ferris, Herdman, & McKinnon, 2011), vividness of future imagery (Stöber, 2000), and concreteness of personal goals (Emmons, 1992).

Yet not all studies have found a negative link between depression and future specificity (e.g., Boelen, Huntjens, & van den Hout, 2014; Robinaugh, Lubin, Babic, & McNally, 2013; Sarkohi, 2011), which makes it difficult at this stage to evaluate the robustness of the effect. And although a recent meta-analysis on prospection in psychopathology found a large effect of reduced future

---

Corresponding Author:
Beau Gamble, School of Psychology, The University of Auckland, Level 2, Building 302, Science Centre, 23 Symonds St., Auckland Central, 1010, New Zealand
E-mail: b.gamble@auckland.ac.nz

specificity in depression, the analysis included only seven studies (Hallford, Austin, Takano, & Raes, 2018). We believe there are many additional studies relevant to this issue and aim to synthesize these findings here. Given that imagining specific, detailed future scenarios is associated with a myriad of benefits, such as helping us to plan, pursue goals, and regulate emotions (Brown, MacLeod, Tata, & Goddard, 2002; Schacter, Addis, & Buckner, 2007; Taylor, Pham, Rivkin, & Armor, 1998), it is worthwhile to quantify precisely how this ability is affected in depression.

Present Study

We aimed to extend the meta-analysis by Hallford et al. (2018) in several important ways. First, we took a broader view of depression, acknowledging that depressive symptoms exist along a continuum of severity throughout the population (Ayuso-Mateos, Nuevo, Verdes, Naidoo, & Chatterji, 2010), and included studies of participants with subthreshold depressive symptoms (often used interchangeably with the term *dysphoria*; e.g., Anderson, Boland, & Garner, 2016; Cropley & MacLeod, 2003; Holmes, Lang, Moulds, & Steele, 2008). Second, we took a wider view of future thinking and included any of the four modes of future thinking from the "taxonomy of prospection" described by Szpunar, Spreng, and Schacter (2014): simulation, prediction, intention, and planning.[1] Essentially, we were interested in any kind of future thinking for which specificity can be (and has been) measured in depression.

And third, we used what we see as a more complete approach to the concept of specificity. Stemming from Tulving's (1972) seminal work on episodic memory and building on earlier studies examining the specificity of autobiographical memory in suicide attempters (Williams & Broadbent, 1986), Williams et al. (1996) conceptualized a specific future thought as one that has a unique spatiotemporal context (i.e., occurring at a particular place and time). However, in addition to spatiotemporal specificity, Tulving (1972) also described the "perceptible properties" of events as being a key component of episodic specificity. Recent research has similarly treated the notion of episodic specificity more broadly than mere spatiotemporal specificity by also considering the level of episodic details within future events, such as information about specific people, objects, actions, feelings, and perceptual details (Addis, Wong, & Schacter, 2008; Jing, Madore, & Schacter, 2016).

In their meta-analysis on episodic specificity, Hallford et al. (2018) included measures of episodic detail but excluded measures of the vividness of future thinking; nonetheless, vividness can be an important indicator of episodic detail (Martin, 2013; Nelson, Moskovitz, & Steiner, 2008). Moreover, vividness captures the clarity and subjective feeling[2] of "preexperiencing" a future event, which are key features of projecting oneself into specific future episodes (Atance & O'Neill, 2001; Tulving, 1985). We therefore considered measures of vividness to capture an important aspect of episodic specificity, in line with Tulving's original definition (Tulving, 1972). The inclusion of studies across a range of depressive symptoms, future thinking, and specificity should not imply that we viewed differences within each of these variables as trivial; on the contrary, we examined the effects of these differences via the moderator analyses described below.

Following a preregistered protocol for meta-analysis (available on the Open Science Framework at https://osf.io/evdkf), we investigated three research questions. First, what is the relationship between depression and specificity (as earlier defined) of future thinking? Second, to what extent can the level of depression be explained by the concurrent variation in specificity of future thinking? While being careful not to infer causality, investigating this question will give a sense of the degree to which depression can be predicted from specificity of future thinking. And third, what moderating variables can explain any heterogeneity in results across studies? We investigated the effects of two categories of moderator variables: those pertaining to participant characteristics and those pertaining to research methodologies. We had specific theoretical predictions about the effect of each moderator.

Moderators in participant characteristics

Research on the specificity of future thinking in depression has spanned a range of demographics and clinical groups. Establishing whether the relationship differs across participant characteristics will help inform who may benefit most from interventions to enhance specificity. We tested the effects of three preregistered moderating variables relating to participants: clinical status of depression, comorbid anxiety, and age.

"Clinical status of depression" refers to the status of patients or participants under investigation. Some studies have examined specificity in individuals with major depressive disorder (MDD) compared with healthy control subjects, whereas other studies have focused on individuals with dysphoria, remitted depression, or even levels of depressive symptoms in those scoring below cutoffs for mild depression. We predicted that reduced specificity of future thinking would be evident at higher levels of depression in the above statuses because (a) subthreshold depressive symptoms are thought not to differ qualitatively with full-blown clinical depression

(Ayuso-Mateos et al., 2010) and (b) reduced specificity of past and future thinking has been found to persist even in individuals with remitted depression (Addis et al., 2016; Brittlebank, Scott, Williams, & Ferrier, 1993; Mackinger, Pachinger, Leibetseder, & Fartacek, 2000).

"Comorbid anxiety" denotes whether participants with depression/dysphoria also had a diagnosis or high level of anxiety. Some studies have indicated enhanced prospective imagery for negative events in those with anxiety (Morina, Deeprose, Pusowski, Schmid, & Holmes, 2011; Stöber, 2000), but the more common result is of reduced specificity (Brown et al., 2013; Kleim, Graham, Fihosy, Stott, & Ehlers, 2014; McNally, Lasko, Macklin, Pitman, & McNally, 1995; McNally, Litz, Prassas, Shin, & Weathers, 1994; Wu, Szpunar, Godovich, Schacter, & Hofmann, 2015). A central feature of anxiety—worry—is thought to be mostly verbal, non-episodic, and nonconcrete (Borkovec & Ray, 1998; Miloyan, Bulley, & Suddendorf, 2016). In addition, boosting the specificity of future imagination through an "episodic specificity induction" has been linked to decreased anxiety toward future events (Jing et al., 2016). Considering these findings, we predicted that future specificity would not be enhanced in those with comorbid anxiety compared with depression alone.

Age was also examined as a moderator, as future specificity is known to vary with age: Older adults tend to generate fewer episodic details and more semantic details than younger adults (Schacter, Gaesser, & Addis, 2013). It has been suggested that age may moderate (increase) the relationship between depression and the overgenerality of memory, as age is associated with more previous depressive episodes and, relatedly, more damage to the hippocampus (King et al., 2010). Whether age influences the relationship between depression and future specificity has, to our knowledge, not been investigated. We predicted that age would increase the strength of the relationship, as future thinking in older depressed adults may suffer the compounding effects of both age and depression.

Moderators in research methodology

In one of the first studies in this area, MacLeod and Cropley (1995) wrote: "What is clear is that there are important distinctions to be made in future-thinking and that any demonstrated relationship between mood disturbance and future-thinking will be affected by the particular measure of future-thinking used" (p. 48). Considering the assortment of methods that have emerged—along with mixed results—their statement is truer now than ever. We assessed four moderators relating to measures of future thinking: emotional valence of simulations, macro- versus microlevel specificity, cue type, and self- versus researcher-rated specificity. We also examined two moderators pertaining to measures of depression: self- versus researcher-rated and categorical versus dimensional.

The "emotional valence of simulations" refers to the positivity or negativity of future simulations.[3] Overly negative future thinking has featured prominently in models of depression since Beck (1976) introduced his cognitive triad and has even been described as "the primary cause" of depression (Roepke & Seligman, 2016, p. 30); it is thus worthwhile to clarify how emotional valence interacts with specificity in this disorder. Some measures such as the Future Event Task (based on the Autobiographical Memory Test, or AMT; Williams et al., 1996) prompt participants with positive-, negative-, or neutral-cue words to elicit future simulations of these valences, but measures of emotional valence may also refer to participant self-ratings following a simulation (e.g., as collected by Addis et al., 2016). A common finding in the memory literature is that the specificity of memory is reduced broadly (across valences) in depression but is particularly affected for positive events (reviewed in van Vreeswijk & de Wilde, 2004). There is evidence for similar effects in future thinking (Dickson & Bates, 2006; Stöber, 2000), but not all findings are consistent. Dysphoric individuals have been found, for example, to generate specific negative events even faster (MacLeod & Cropley, 1995) and more vividly (Holmes et al., 2008) than healthy control subjects. As episodic memory and simulation rely on similar neurocognitive processes (reviewed in Addis, 2018; Schacter et al., 2012), we predicted future thinking in depression would reveal a pattern similar to that evident for memory; that is, it would be less specific for all emotional valences, but the deficit would be strongest for positive future thinking.

"Macro- versus microspecificity" distinguishes the event level at which specificity is measured.[4] Following the approach of Hach, Tippett, and Addis (2016), *macro* refers to the spatiotemporal specificity of the future event itself, for example, whether the event is localized to a time and place. Such measures are typically categorical; events are scored as specific or not (e.g., Williams et al., 1996). *Micro* refers to measures of specificity within future simulations, such as the number of episodic details (King et al., 2011), self-reported level of detail (e.g., Addis et al., 2016), and vividness/clarity of future imagery (e.g., Stöber, 2000). Micro measures are often continuous (e.g., when the outcome variable is the number of details). We are not aware of any thorough attempt to examine how macro- and micro-level specificity relate differentially to depression. We

predicted that both levels would be impaired; in addition, considering that the categorization of continuous variables reduces the statistical power to detect true effects (Hunter & Schmidt, 1990), we predicted macro measures would evidence smaller reductions than micro measures.

"Cue type" refers to the prompt used to elicit a future simulation or event. For instance, the AMT typically uses single words as cues (e.g., "laughing," "friendly," "bread"; e.g., Williams et al., 1996), whereas other studies have used event cues (e.g., "New Year's Eve," "an accident," "an election"; e.g., Addis et al., 2016). Event cues are thought to be more supportive than single-word cues for retrieving and generating specific events (Addis et al., 2016); hence, we predicted that reduced specificity in depression would be less pronounced in studies using the more supportive event cues compared with single-word cues.[5]

"Self- versus researcher-rated specificity" describes whether measures of specificity were attained through participant self-report or researcher coding/count. We predicted the relationship between depression and specificity would be stronger (more negative) for self-reported than researcher-rated outcomes given a tendency for depressed individuals to underestimate their performance on cognitive tasks (Farrin, Hull, Unwin, Wykes, & David, 2003).

"Self- versus researcher-rated depression" refers to whether depression/dysphoria was quantified via self-report questionnaires, such as the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, & Mock, 1961) or Centre For Epidemiological Studies Depression Scale (Radloff, 1977), or evaluated by a clinician or researcher using, for example, the Hamilton Rating Scale for Depression (Hamilton, 1960) or Structured Clinical Interview for *DSM* Disorders (First, Spitzer, Williams, & Gibbon, 1995). Reductions in specificity have been found in both self-rated and clinically diagnosed depressed samples, and we are not aware of any theoretical basis to suggest the effect is heightened for either type of measure; thus, we predicted no effect of this moderator.

"Categorical versus dimensional depression" denotes whether researchers used a categorical (group) or dimensional (continuous) design to assess depression. A recent meta-analysis on interpretation biases in depression found significantly stronger effects in studies with dimensional than categorical designs (Everaert, Podina, & Koster, 2017). This finding makes sense statistically—dichotomizing continuous variables can lead to an underestimation of the strength of relationships (Hunter & Schmidt, 1990). We thus predicted the relation between depression and future specificity would be stronger for studies that were dimensional rather than categorical.

Method

We designed, preregistered, and reported the results of the meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Moher et al., 2015). Figure 1 depicts the major steps of the meta-analysis; any deviations from the preregistration are noted in the text and endnotes.

Inclusion criteria

The criteria for including a study in the meta-analysis are listed in Figure 1 according to the PICOS categories (population, intervention, comparison, outcomes, study design; Moher et al., 2015).[6] One requirement was that a study measured some aspect of specificity in future thinking. As described a priori, we included measures that directly assessed one or both key components of episodic specificity (Tulving, 1972): spatiotemporal specificity (usually reported as a proportion of all events that are specific in time and place) and measures of perceptible properties of events (e.g., number of details, vividness). Moreover, measures of constructs that are close semantic associates of *specific* (i.e., lay within two degrees of separation from *specific* in the Oxford English Thesaurus) were also included.[7]

A common measure of future thinking in psychopathology has been the Future Thinking Task (FTT), which assesses verbal fluency for future events (MacLeod & Byrne, 1996).[8] As planned, strict measures of future-event fluency were not included in the meta-analysis for two reasons. First, even with instructions for participants to be as specific as possible, one of the only studies to code responses for specificity found that a substantial proportion of named events were in fact general (27.5% and 14% of events generated by dysphoric and control subjects, respectively; MacLeod & Cropley, 1995). This observation suggests that without coding of events for specificity, the FTT score does not necessarily capture specificity. The second reason was theoretical—future-event fluency tasks have been proposed to largely measure the ability to access *abstract* information about the future (D'Argembeau, Ortoleva, Jumentier, & Van der Linden, 2010). Thus, FTT data on the number of events generated (unless coded for specificity) were not included.

Literature search and coding

To be indexed, studies had to mention at least one search term relating to each of three key variables: depression, future thinking, and specificity (see Table 1 for a list of all search terms). We searched the databases PsychINFO, Scopus, PubMed, ScienceDirect, Web
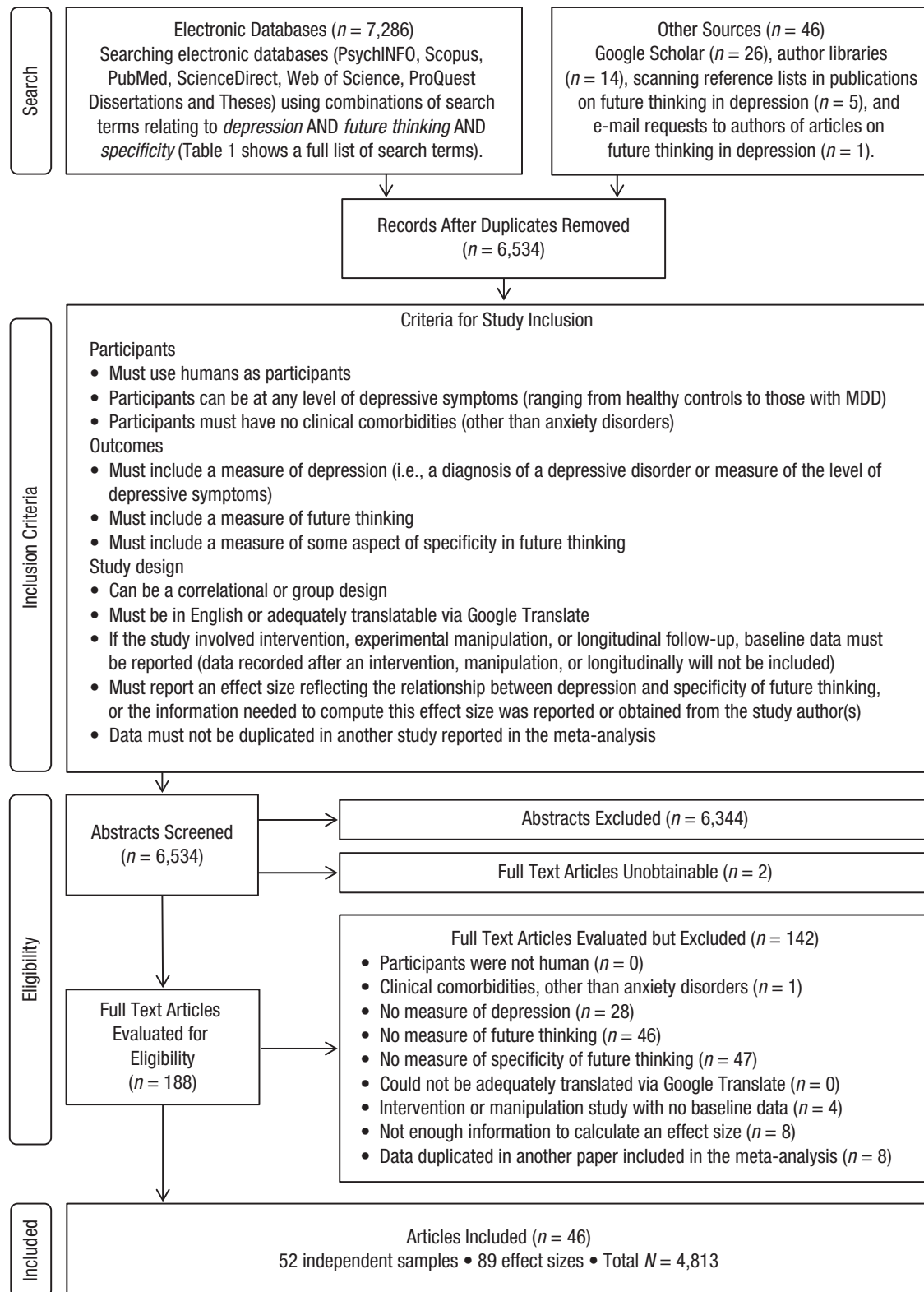
Search

Electronic Databases (*n* = 7,286)
Searching electronic databases (PsychINFO, Scopus, PubMed, ScienceDirect, Web of Science, ProQuest Dissertations and Theses) using combinations of search terms relating to *depression* AND *future thinking* AND *specificity* (Table 1 shows a full list of search terms).

Other Sources (*n* = 46)
Google Scholar (*n* = 26), author libraries (*n* = 14), scanning reference lists in publications on future thinking in depression (*n* = 5), and e-mail requests to authors of articles on future thinking in depression (*n* = 1).

Records After Duplicates Removed
(*n* = 6,534)

Inclusion Criteria

Criteria for Study Inclusion

Participants
- Must use humans as participants
- Participants can be at any level of depressive symptoms (ranging from healthy controls to those with MDD)
- Participants must have no clinical comorbidities (other than anxiety disorders)

Outcomes
- Must include a measure of depression (i.e., a diagnosis of a depressive disorder or measure of the level of depressive symptoms)
- Must include a measure of future thinking
- Must include a measure of some aspect of specificity in future thinking

Study design
- Can be a correlational or group design
- Must be in English or adequately translatable via Google Translate
- If the study involved intervention, experimental manipulation, or longitudinal follow-up, baseline data must be reported (data recorded after an intervention, manipulation, or longitudinally will not be included)
- Must report an effect size reflecting the relationship between depression and specificity of future thinking, or the information needed to compute this effect size was reported or obtained from the study author(s)
- Data must not be duplicated in another study reported in the meta-analysis

Eligibility

Abstracts Screened
(*n* = 6,534)

Abstracts Excluded (*n* = 6,344)

Full Text Articles Unobtainable (*n* = 2)

Full Text Articles Evaluated for Eligibility
(*n* = 188)

Full Text Articles Evaluated but Excluded (*n* = 142)
- Participants were not human (*n* = 0)
- Clinical comorbidities, other than anxiety disorders (*n* = 1)
- No measure of depression (*n* = 28)
- No measure of future thinking (*n* = 46)
- No measure of specificity of future thinking (*n* = 47)
- Could not be adequately translated via Google Translate (*n* = 0)
- Intervention or manipulation study with no baseline data (*n* = 4)
- Not enough information to calculate an effect size (*n* = 8)
- Data duplicated in another paper included in the meta-analysis (*n* = 8)

Included

Articles Included (*n* = 46)
52 independent samples • 89 effect sizes • Total *N* = 4,813

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the literature search and study coding. The last inclusion criterion (i.e., that data were not duplicated) was added after the pre-registration because it was not anticipated during planning; the phrasing of some criteria has also been amended for clarity. *n* = number of studies; *N* = number of participants across all included studies.

Table 1. Literature Search Terms

Depression	Future thinking	Specificity
depress*	future thinking	more specific
dysphor*	future thought*	less specific
dysthymi*	future-directed thinking	number of specific
	future-directed thought*	greater detail*
	prospective	reduced detail*
	episodic simulation*	increased detail*
	future simulation*	number of detail*
	future event*	internal detail*
	future imag*	external detail*
	autobiographical plan*	episodic detail*
	personal project*	semantic detail*
	personal striving*	vivid*
	scene construction	enhanced imag*
	constructive daydream*	reduced imag*
		overgeneral*
		concrete*

Note: To be indexed, studies had to mention at least one search term from each column (i.e., *depression* AND *future thinking* AND *specificity*). The search terms deviate slightly from those listed in the preregistration, which initially yielded an impractically large number of results (e.g., the first search in Scopus returned 40,984 articles). To limit irrelevant results, broad terms such as *goal, specificity, more detail*, and *less detail* were removed, and *prospect* was changed to *prospective* (which rendered the terms *prospective imag** and *prospective cognition** redundant, so these were also removed). The wide array of search terms reflects the plethora of constructs researchers have used to assess aspects of future thinking.

of Science, and ProQuest Dissertations and Theses on December 1, 2017, and again on January 14, 2019, while this manuscript was under review. Google Scholar was searched on December 13, 2017 (and again on January 14, 2019) using modified search terms because of the 256-character search limit. No study design, date, or language limits were imposed on the searches.[9] After the original search, e-mails were sent to 40 authors of articles on future thinking in depression requesting unpublished or in-press data, with a deadline for responses set to January 26, 2018. We also scanned reference lists of included studies and related reviews (*n* = 20) and searched our personal files to capture as many relevant studies as possible.

A spreadsheet containing all the search results (*n* = 7,332) is available at https://osf.io/x3jy6. B. Gamble screened all titles and abstracts and then removed duplicates (*n* = 798) and any studies clearly not meeting inclusion criteria (*n* = 6,344). B. Gamble then reviewed the full reports of titles appearing to meet inclusion criteria or where there was any uncertainty (*n* = 188).[10] For any study excluded, the primary inclusion criteria not met was recorded. We sought additional information from study authors where necessary to resolve questions about eligibility or where data

were insufficient to calculate an effect size. The full-text article for each study deemed eligible by B. Gamble (*n* = 46) as well as any ambiguous cases (*n* = 1) were further reviewed by D. Moreau and D. R. Addis for confirmation of inclusion. Agreement between all three authors was unanimous in all cases (i.e., 100%). We identified 46 articles meeting inclusion criteria, with 52 independent samples and 89 effect sizes (*N* = 4,813). Next, B. Gamble extracted reference information, methodological characteristics, and results from eligible studies. D. Moreau reviewed the data extracted by B. Gamble and compared it to the full-text articles to confirm that the extracted data were accurate; no discrepancies were found. The final data file used for analysis, which contains the extracted data, describes all variables for which data were sought, and includes notes on any difficult decisions during coding, is openly available at https://osf.io/a6q5y.

Effect sizes and moderator variables

The measure of effect size was the correlation between level of depressive symptoms and the specificity of future thinking. We took a correlational rather than categorical approach because, as mentioned, depressive symptoms are thought to occur along a continuum of severity throughout the population (Ayuso-Mateos et al., 2010). For studies in which only group-level comparisons were reported (e.g., patients with MDD vs. healthy control subjects), standardized mean differences (Cohen's *d*) were converted to biserial correlations (Becker, 1986; Hunter & Schmidt, 1990).[11]

The coding of moderator variables was mostly straightforward and followed the preregistration. For the moderator variable cue type, an additional (unplanned) subgroup was needed to accommodate studies that did not fit into the categories of single-word or event cues. A new subgroup, "open," was created for cues that were more open-ended, such as those used in sentence-completion tasks (e.g., "I can imagine that, shortly, I . . ."; Boelen et al., 2014). Thirty-eight effect sizes fell into the open category, so we considered it an important addition despite deviating from the preregistration. For the moderator variable emotional valence, a new subgroup, "combined," was added to include studies for which only results of positive and negative conditions together were reported.

As a rule of thumb, at least four independent samples are recommended per subgroup for metaregression (Fu et al., 2011). For two moderator variables, some subgroups did not meet this threshold: Three of the six subgroups of the moderator clinical status of depression contained effect sizes from just two independent samples, and the two subgroups of the moderator comorbid

anxiety contained effect sizes from only three independent samples.[12] Rather than exclude subgroups with a small number of cases, we ran the moderator analyses as planned but interpreted results with caution, considering possibly insufficient power to detect true effects in those subgroups.

Meta-analytic procedure

We used random-effects meta-analysis modeling[13] with restricted maximum likelihood[14] to estimate overall effects and the heterogeneity across included studies. We also used mixed-effects meta-analysis modelling to test whether differences in the strength of effect sizes across studies could be explained by the moderator variables. Analyses were run in the R software environment (Version 3.6.0; R Core Team, 2019) using the *metafor* package (Viechtbauer, 2010), and our R script is available online (https://osf.io/35kzx). Some studies included multiple measures of future thinking or depression within the same sample, so we accounted for dependency by adding random effects corresponding to each independent sample (Viechtbauer, 2010). We calculated 95% confidence intervals (CIs) for the overall effect size and inferred confidence in the cumulative estimate from a combination of the magnitude and precision of the effect size and risks of publication and reporting bias. Biases were assessed using a combination of trim-and-fill and *p*-curve analysis, published versus unpublished study comparisons, and examination of study quality. B. Gamble and D. Moreau assessed study quality[15] using the 18-item Checklist for Measuring Quality (Everaert et al., 2017) adapted by Downs and Black (1998) to exclude items related to interventions, making it suitable for the current meta-analysis.

Outliers were predefined as correlations whose residuals had z scores > 3. No studies met this threshold (see Fig. S1 in the Supplemental Material available online), and so none were excluded from primary analyses. Noting that some effect sizes deviated greatly from the mean (see Fig. 2), we also explored an alternate measure of outliers, Cook's distance (D_i), which indicates the relative influence of each effect size on the summary estimate. A standard rule of thumb is that D_i values greater than three times the mean D_i may be potential outliers. Seven effect sizes exceeded this threshold (see Fig. S2 in the Supplemental Material), and so we ran exploratory analyses after their exclusion as a sensitivity check. No substantial changes occurred to the summary estimate or results of moderator analyses, except for the moderator categorical versus dimensional depression, which shifted from borderline statistically significant to nonsignificant after the removal of outliers; this point is addressed further in the Discussion section. These secondary analyses are

available online (https://osf.io/cvqg4), but we focus here on the initial analyses run with all effect sizes.

Results

The studies encompassed a range of participants, from symptom-free to clinically depressed, and a multiplicity of future-thinking tasks—across the 46 included articles, researchers used 26 different measures to capture some aspect of specificity in future thinking (study characteristics for each effect size are described in Table S1 in the Supplemental Material). Figure 2 shows that almost two thirds (57 of 89) of correlations between depression and future specificity were negative; that is, higher levels of depression were associated with reduced specificity of future thinking. The meta-analytic average correlation was $r = -.13$, 95% CI = [−.20, −.05], $p < .001$, indicating future specificity could explain 1.6%, 95% CI = [0.3%, 4.0%], of the variance in depression, leaving 98.4% of the variance unexplained. There was a high degree of heterogeneity across effect sizes, as might be expected from the diversity of future thinking tasks. The I^2 statistic, which indicates the percentage of between-studies variability in effect sizes due to heterogeneity rather than random error, was $I^2 = 88.3$ for the overall model. Given the high level of heterogeneity, we characterized the underlying distribution of effect sizes via mixture modeling following the procedure described in Moreau and Corballis (2019). We estimated that the distribution of effect sizes was well characterized by a single-component distribution, suggesting no manifest departure from normality (see Fig. S8 in the Supplemental Material). In the analyses reported next, we investigated whether some of the heterogeneity across studies could be explained by the moderator variables.

Moderator analyses of participant characteristics

The effect of clinical status of depression was not significant, $Q(5) = 9.19$, $p = .10$. As predicted, the correlation between specificity and depression was negative and significant in samples spanning nondepressed to clinically depressed participants, $r = -.23$, 95% CI = [−0.40, −0.05], $p = .010$, $k = 16$, and nondepressed to dysphoric participants, $r = -.12$, 95% CI = [−.20, −.04], $p = .005$, $k = 61$.[16] We predicted the effect would hold across all statuses of depression, but this was not the case: The effect was nonsignificant in samples spanning nondepressed to remitted participants, $r = .10$, 95% CI = [−.17, 36], $p = .48$, $k = 2$, and samples including only (i.e., with no comparison group) participants who were nondepressed, $r = -.13$, 95% CI = [−.33, .07], $p = .20$, $k = 5$; clinically depressed, $r = .07$, 95% CI = [−.31,
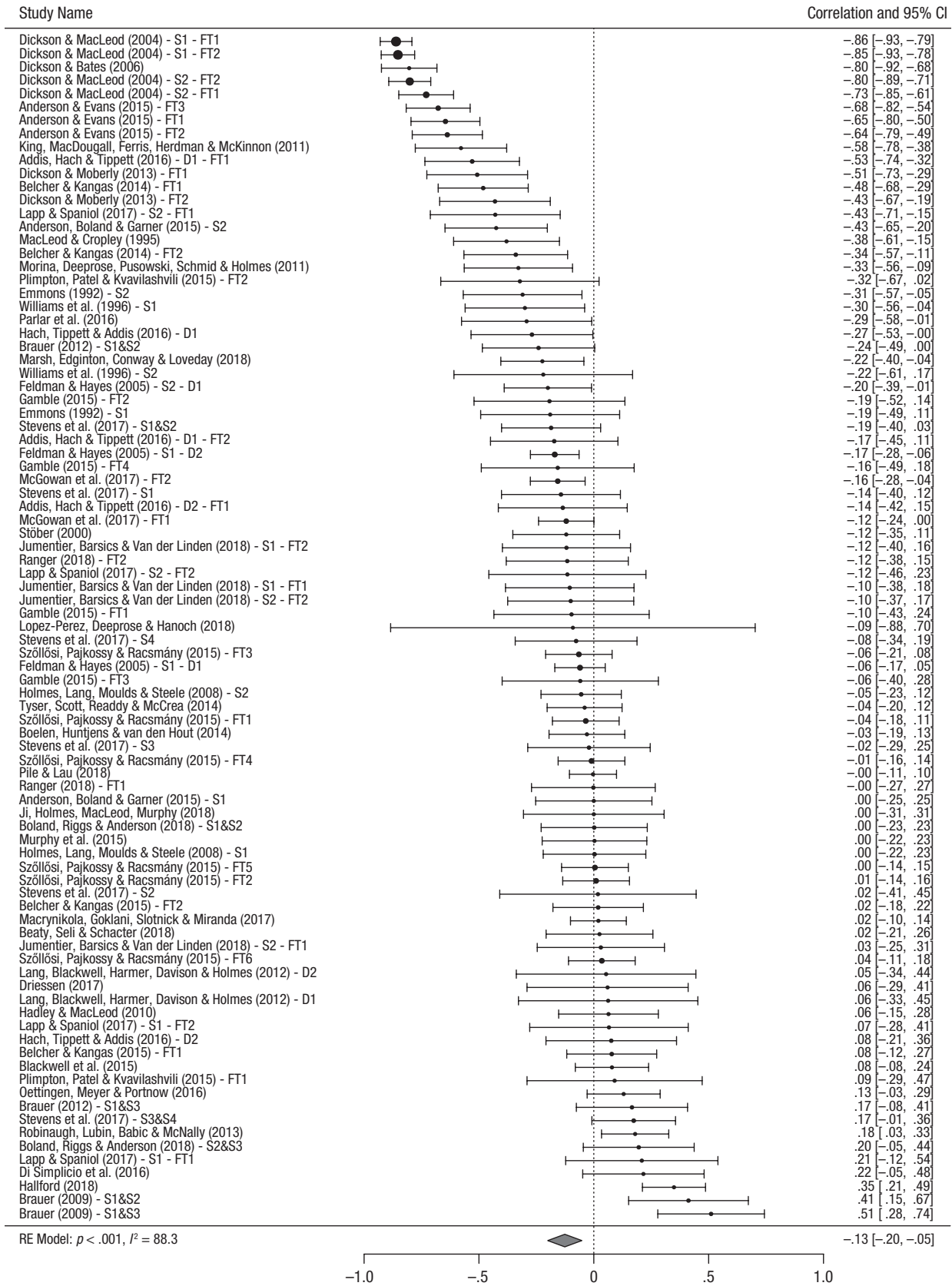
Study Name .... **Correlation and 95% CI**

Study Name	Correlation and 95% CI
Dickson & MacLeod (2004) - S1 - FT1	−.86 [−.93, −.79]
Dickson & MacLeod (2004) - S1 - FT2	−.85 [−.93, −.78]
Dickson & Bates (2006)	−.80 [−.92, −.68]
Dickson & MacLeod (2004) - S2 - FT2	−.80 [−.89, −.71]
Dickson & MacLeod (2004) - S2 - FT1	−.73 [−.85, −.61]
Anderson & Evans (2015) - FT3	−.68 [−.82, −.54]
Anderson & Evans (2015) - FT1	−.65 [−.80, −.50]
Anderson & Evans (2015) - FT2	−.64 [−.79, −.49]
King, MacDougall, Ferris, Herdman & McKinnon (2011)	−.58 [−.78, −.38]
Addis, Hach & Tippett (2016) - D1 - FT1	−.53 [−.74, −.32]
Dickson & Moberly (2013) - FT1	−.51 [−.73, −.29]
Belcher & Kangas (2014) - FT1	−.48 [−.68, −.29]
Dickson & Moberly (2013) - FT2	−.43 [−.67, −.19]
Lapp & Spaniol (2017) - S2 - FT1	−.43 [−.71, −.15]
Anderson, Boland & Garner (2015) - S2	−.43 [−.65, −.20]
MacLeod & Cropley (1995)	−.38 [−.61, −.15]
Belcher & Kangas (2014) - FT2	−.34 [−.57, −.11]
Morina, Deeprose, Pusowski, Schmid & Holmes (2011)	−.33 [−.56, −.09]
Plimpton, Patel & Kvavilashvili (2015) - FT2	−.32 [−.67, .02]
Emmons (1992) - S2	−.31 [−.57, −.05]
Williams et al. (1996) - S1	−.30 [−.56, −.04]
Parlar et al. (2016)	−.29 [−.58, −.01]
Hach, Tippett & Addis (2016) - D1	−.27 [−.53, −.00]
Brauer (2012) - S1&S2	−.24 [−.49, −.00]
Marsh, Edginton, Conway & Loveday (2018)	−.22 [−.40, −.04]
Williams et al. (1996) - S2	−.22 [−.61, .17]
Feldman & Hayes (2005) - S2 - D1	−.20 [−.39, −.01]
Gamble (2015) - FT2	−.19 [−.52, .14]
Emmons (1992) - S1	−.19 [−.49, .11]
Stevens et al. (2017) - S1&S2	−.19 [−.40, .03]
Addis, Hach & Tippett (2016) - D1 - FT2	−.17 [−.45, .11]
Feldman & Hayes (2005) - S1 - D2	−.17 [−.28, −.06]
Gamble (2015) - FT4	−.16 [−.49, .18]
McGowan et al. (2017) - FT2	−.16 [−.28, −.04]
Stevens et al. (2017) - S1	−.14 [−.40, .12]
Addis, Hach & Tippett (2016) - D2 - FT1	−.14 [−.42, .15]
McGowan et al. (2017) - FT1	−.12 [−.24, .00]
Stöber (2000)	−.12 [−.35, .11]
Jumentier, Barsics & Van der Linden (2018) - S1 - FT2	−.12 [−.40, .16]
Ranger (2018) - FT2	−.12 [−.38, .15]
Lapp & Spaniol (2017) - S2 - FT2	−.12 [−.46, .23]
Jumentier, Barsics & Van der Linden (2018) - S1 - FT1	−.10 [−.38, .18]
Jumentier, Barsics & Van der Linden (2018) - S2 - FT2	−.10 [−.37, .17]
Gamble (2015) - FT1	−.10 [−.43, .24]
Lopez-Perez, Deeprose & Hanoch (2018)	−.09 [−.88, .70]
Stevens et al. (2017) - S4	−.08 [−.34, .19]
Szőllősi, Pajkossy & Racsmány (2015) - FT3	−.06 [−.21, .08]
Feldman & Hayes (2005) - S1 - D1	−.06 [−.17, .05]
Gamble (2015) - FT3	−.06 [−.40, .28]
Holmes, Lang, Moulds & Steele (2008) - S2	−.05 [−.23, .12]
Tyser, Scott, Readdy & McCrea (2014)	−.04 [−.20, .12]
Szőllősi, Pajkossy & Racsmány (2015) - FT1	−.04 [−.18, .11]
Boelen, Huntjens & van den Hout (2014)	−.03 [−.19, .13]
Stevens et al. (2017) - S3	−.02 [−.29, .25]
Szőllősi, Pajkossy & Racsmány (2015) - FT4	−.01 [−.16, .14]
Pile & Lau (2018)	−.00 [−.11, .10]
Ranger (2018) - FT1	−.00 [−.27, .27]
Anderson, Boland & Garner (2015) - S1	.00 [−.25, .25]
Ji, Holmes, MacLeod, Murphy (2018)	.00 [−.31, .31]
Boland, Riggs & Anderson (2018) - S1&S2	.00 [−.23, .23]
Murphy et al. (2015)	.00 [−.22, .23]
Holmes, Lang, Moulds & Steele (2008) - S1	.00 [−.22, .23]
Szőllősi, Pajkossy & Racsmány (2015) - FT5	.00 [−.14, .15]
Szőllősi, Pajkossy & Racsmány (2015) - FT2	.01 [−.14, .16]
Stevens et al. (2017) - S2	.02 [−.41, .45]
Belcher & Kangas (2015) - FT2	.02 [−.18, .22]
Macrynikola, Goklani, Slotnick & Miranda (2017)	.02 [−.10, .14]
Beaty, Seli & Schacter (2018)	.02 [−.21, .26]
Jumentier, Barsics & Van der Linden (2018) - S2 - FT1	.03 [−.25, .31]
Szőllősi, Pajkossy & Racsmány (2015) - FT6	.04 [−.11, .18]
Lang, Blackwell, Harmer, Davison & Holmes (2012) - D2	.05 [−.34, .44]
Driessen (2017)	.06 [−.29, .41]
Lang, Blackwell, Harmer, Davison & Holmes (2012) - D1	.06 [−.33, .45]
Hadley & MacLeod (2010)	.06 [−.15, .28]
Lapp & Spaniol (2017) - S1 - FT2	.07 [−.28, .41]
Hach, Tippett & Addis (2016) - D2	.08 [−.21, .36]
Belcher & Kangas (2015) - FT1	.08 [−.12, .27]
Blackwell et al. (2015)	.08 [−.08, .24]
Plimpton, Patel & Kvavilashvili (2015) - FT1	.09 [−.29, .47]
Oettingen, Meyer & Portnow (2016)	.13 [−.03, .29]
Brauer (2012) - S1&S3	.17 [−.08, .41]
Stevens et al. (2017) - S3&S4	.17 [−.01, .36]
Robinaugh, Lubin, Babic & McNally (2013)	.18 [.03, .33]
Boland, Riggs & Anderson (2018) - S2&S3	.20 [−.05, .44]
Lapp & Spaniol (2017) - S1 - FT1	.21 [−.12, .54]
Di Simplicio et al. (2016)	.22 [−.05, .48]
Hallford (2018)	.35 [.21, .49]
Brauer (2009) - S1&S2	.41 [.15, .67]
Brauer (2009) - S1&S3	.51 [.28, .74]
RE Model: $p < .001$, $I^2 = 88.3$	−.13 [−.20, −.05]

−1.0   −.5   0   .5   1.0

Fig. 2. Correlations (*r*) between levels of depression and specificity of future thinking. Correlations (dots) are shown for all effects in the meta-analysis. The size of each dot reflects the weight given to the observed effect during model fitting. Error bars indicate 95% confidence intervals (CIs). The diamond at the bottom shows the meta-analytically weighted mean correlation (with 95% CI). The dotted vertical line represents an effect size of 0. Multiple measures were adjusted for dependency (see Meta-Analytic Procedure in the Method section). Multiple independent samples within studies are reported separately (S1, S2, etc.), as are multiple measures of future thinking (FT1, FT2, etc.) and depression (D1, D2, etc.).
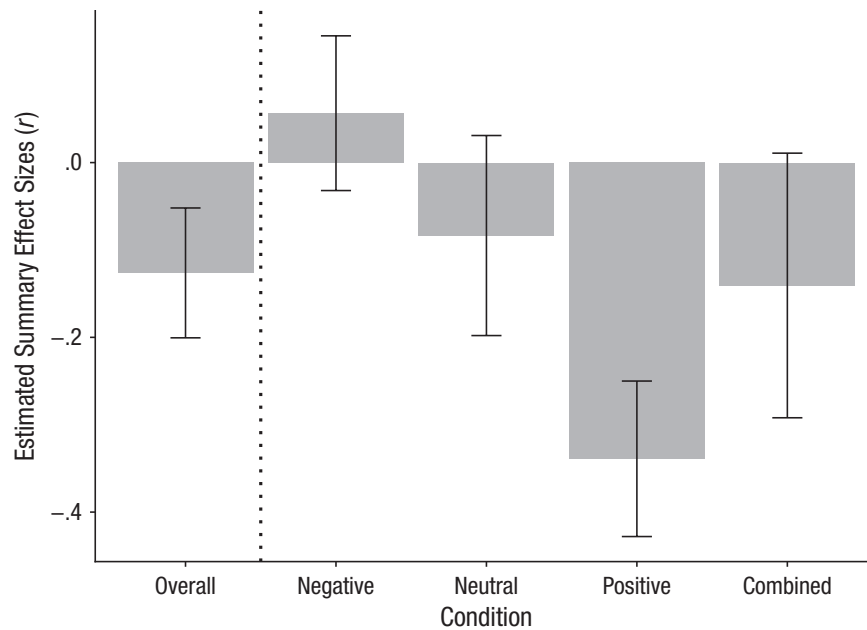
Fig. 3. Relationship between estimated summary effect sizes and each subgroup of emotional valence. The combined condition includes studies for which results of only positive and negative conditions together were reported. Error bars indicate 95% confidence intervals.

.44], $p = .72$, $k = 3$; or dysphoric, $r = -.18$, 95% CI = [.45, .08], $p = .17$, $k = 2$. Given the paucity of effect sizes in these latter four subgroups, and hence the possibility of insufficient power, the null results should be interpreted with caution.

The effect of comorbid anxiety was not significant, $Q(1) = 3.61$, $p = .057$, in line with our prediction. The relationship between specificity and depression was negative and significant in samples without comorbid anxiety, $r = -.35$, 95% CI = [−.65, −.05], $p = .023$, $k = 6$, and negative but nonsignificant in samples with comorbid anxiety, $r = -.26$, 95% CI = [−.57, .04], $p = .087$, $k = 6$. Again, the subgroups contained few effect sizes, possibly hampering our ability to detect a true difference.

The effect of age was not significant, $Q(1) = 0.07$, $b = -0.001$, 95% CI = [−0.006, 0.005], $p = .79$, which diverged from our prediction of a stronger relationship between specificity and depression in older individuals. It is worth noting that samples were heavily skewed toward younger adults, with a mean age (weighted by n) of 25.1 years, and that only five independent samples had a mean age greater than 50 years.

Moderator analyses of research methodology

The effect of emotional valence of simulations was highly significant, $Q(3) = 425.04$, $p < .001$ (see Fig. 3), and partly aligned with our hypotheses. As predicted,

the relationship between depression and specificity was strongest (most negative) in conditions with positive cues or in which future thinking was rated positively, $r = -.34$, 95% CI = [−.43, −.25], $p < .001$, $k = 42$. We hypothesized that the effect would be weaker but still significant across all emotional valences, but this was not the case: For future thinking that was neutral in valence, the effect was weaker and did not reach significance, $r = -.08$, 95% CI = [−.20, .03], $p = .15$, $k = 26$, and for negative future thinking, the effect was close to zero, $r = .06$, 95% CI = [−.03, .15], $p = .21$, $k = 52$. We ran pairwise comparisons on these three subgroups of emotional valence with Bonferroni-adjusted p values to control the family-wise error rate. The difference between positive and neutral subgroups was significant, $r_{diff} = .26$, 95% CI = [.13, .38], $p < .001$, as was the difference between positive and negative subgroups, $r_{diff} = .40$, 95% CI = [.36, .43], $p < .001$. The difference between neutral and negative subgroups was not significant, $r_{diff} = .14$, 95% CI = [.02, .26], $p = .071$. In cases in which studies reported values only for positive and negative conditions combined, the correlation between depression and specificity was nonsignificant, $r = -.14$, 95% CI = [.01, −.29], $p = .067$, $k = 13$. See Figures S3 to S6 in the Supplemental Material for forest plots that show the effect sizes within each subgroup.

The effect of macro- versus microlevel specificity was not significant, $Q(1) = 1.30$, $p = .25$, diverging from our prediction of a stronger effect for micro measures. The

relationship between depression and specificity was negative and significant for both macro, $r = -.15$, 95% CI = [−.23, −.06], $p < .001$, $k = 38$, and micro, $r = -.11$, 95% CI = [−.19, −.03], $p = .005$, $k = 51$, measures of specificity.

The effect of cue type was not significant, $Q(2) = 2.54$, $p = .28$, diverging from our prediction of a stronger effect for single-word than event cues. Although the difference between subgroups was not significant, the correlation between depression and specificity was largest for studies using single-word cues, $r = -.21$, 95% CI = [−.35, −.07], $p = .003$, $k = 27$; the effect was also significant for open-ended cues, $r = -.11$, 95% CI = [−.22, −.01], $p = .037$, $k = 38$, but not for event cues, $r = -.05$, 95% CI = [−.20, .10], $p = .53$, $k = 27$.

The effect of self- versus researcher-rated specificity was not significant, $Q(1) = 3.22$, $p = .07$, diverging from our prediction of a stronger effect for self-rated measures of specificity. The relationship between depression and specificity was negative and significant when researchers coded participants' responses, $r = -.17$, 95% CI = [−.26, −.08], $p < .001$, $k = 47$. The effect was also negative, but nonsignificant, when participants rated their own thoughts, $r = -.07$, 95% CI = [−.17, .02], $p = .14$, $k = 42$.

The effect of self- versus- researcher-rated depression was also not significant, $Q(1) = 1.41$, $p = .23$, in line with our prediction. The relationship between depression and specificity was negative and significant when depression was diagnosed/scored by a clinician or researcher, $r = -.25$, 95% CI = [−.46, −.03], $p = .027$, $k = 7$, and when assessed via participant self-report, $r = -.11$, 95% CI = [−.18, −.03], $p = .006$, $k = 79$.

The effect of categorical versus dimensional measures of depression was significant, $Q(1) = 3.90$, $p = .048$, but in the nonpredicted direction. The relationship between depression and specificity was negative and significant for both subgroups but stronger (more negative) for categorical, $r = -.19$, 95% CI = [−.29, −.09], $p < .001$, $k = 35$, than dimensional, $r = -.09$, 95% CI = [−.17, −.01], $p = .035$, $k = 54$, measures. This was the only moderator for which the removal of outliers (described in the Method section) altered the interpretation of results; the effect shifted from significant to nonsignificant, $Q(1) = 3.31$, $p = .07$, after removing outliers.

Exploratory moderator analyses

In addition to planned, preregistered analyses, we also explored the effects of two moderator variables not identified in the preregistration: sex of participants and mode of future thinking. We made no prediction about the moderating influence of sex but were motivated to

explore its effects given the previously reported sex differences in the detail and vividness of autobiographical memories (Grysman & Hudson, 2013). The effect of sex was strong and statistically significant, $Q(1) = 5.05$, $b = 0.39$, 95% CI = [0.05, 0.74], $p = .025$, indicating the correlation between depression and specificity was weaker (less negative) in samples with a higher percentage of females. In general, samples were skewed toward more females—across the 46 studies, 68% of all participants were females, and only three samples had more males than females.

The moderator variable mode of future thinking was added to explore differences across simulation, intention, prediction, and planning modes of prospection (Szpunar, Spreng, & Schacter, 2014). For example, some measures elicited future-event simulations (e.g., the Adapted Autobiographical Interview in King et al., 2011), whereas others related explicitly to intentions (e.g., the Personal Strivings Listing in Emmons, 1992) and/or planning (e.g., the Measure for Eliciting Positive Future Goals and Plans used by Hadley & MacLeod, 2010). None of the included studies was categorized as being the prediction mode of future thinking.[17] We expected the relationship between specificity and depression to be stronger for intentions than simulations given that goals are generally positive and having noted the strong effect for positive future thinking described above. Depression has also been closely linked to goal dysregulation in previous studies (Street, 2002). However, the effect of the mode of future thinking was not significant, $Q(1) = 0.46$, $p = .80$, with the relationship between depression and specificity being of a similar magnitude for simulation, $r = -.13$, 95% CI = [−.21, −.05], $p = .002$, $k = 69$; intention, $r = -.10$, 95% CI = [−.24, −.03], $p = .134$, $k = 14$; and planning modes, $r = -.13$, 95% CI = [−.28, .02], $p = .090$, $k = 6$.

Assessment of bias

We used several techniques to examine publication and reporting biases across the included studies. First, we examined whether the meta-analysis showed evidence of "small-study effects"—wherein smaller studies often show different, stronger effects than larger studies, possibly reflecting publication bias (Schwarzer, Carpenter, & Rücker, 2015). To this end we inspected a funnel plot of the relationship between effect size and standard error. If a meta-analysis is free from small-study effects, effect sizes derived from larger samples (and thus with smaller standard errors) are expected to cluster around the mean, whereas effect sizes derived from smaller samples (and thus with larger standard errors) should be broadly dispersed and distributed symmetrically around the mean, forming a funnel-like shape. The
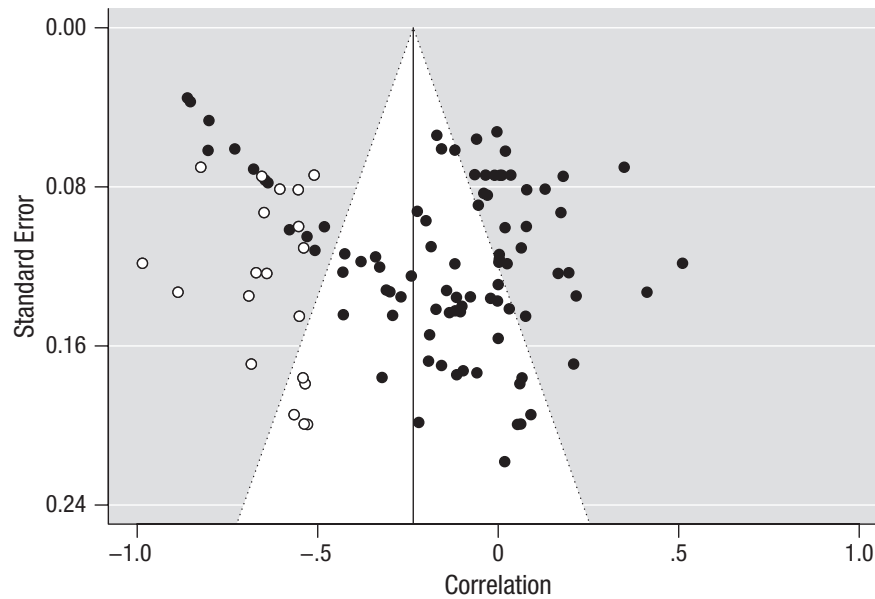
Fig. 4. Funnel plot of observed effect sizes (black circles) and those estimated by trim-and-fill analysis to be missing from the meta-analysis (white circles). Each dot represents an individual effect size and is positioned according to its correlation coefficient (*x*-axis) and standard error (*y*-axis). The vertical line represents the overall estimated effect if corrected by the trim-and-fill analysis for publication bias (e.g., *r* would equal −.24 rather than the −.13 from the main model). The white triangle represents a 95% confidence interval around the corrected estimated effect.

funnel plot for this meta-analysis clearly deviated from an unbiased model (see Fig. 4). Oddly, a number of effect sizes with small standard errors were dispersed far from the mean, particularly toward the left side of the plot; in other words, a few studies with apparent high precision showed very strong negative correlations between depression and future specificity. This suggested the meta-analysis may have missed other correlations that were also stronger (more negative) than the mean.

A trim-and-fill analysis from Duval and Tweedie (2000a, 2000b) supported this interpretation, estimating that 15 effect sizes were missing from the left side of the funnel plot (shown as white circles in Fig. 4), and that the inclusion of the missing effect sizes in the random-effects model would increase the magnitude of the summary estimate to *r* = −.24, 95% CI = [−0.30, −0.17], *p* < .001. This observation is the reverse of what would be expected from typical publication bias, in which missing unpublished studies are those with weak or null effects.

Second, to test for inflation of the effect in the published literature relative to the true effect, we compared the magnitude of effect sizes in published versus unpublished studies. A moderator analysis showed that the effect of publication status was significant, *Q*(1) = 5.68, *p* = .017, indicating stronger (more negative) correlations

between depression and specificity in published studies, *r* = −.15, 95% CI = [−.22, −.07], *p* < .001, *k* = 76, versus unpublished studies, *r* = .06, 95% CI = [−.11, .23], *p* = .48, *k* = 13). In contrast to inspection of the funnel plot, this analysis indicated typical publication bias; that is, the true correlation between depression and specificity may be smaller than the summary estimate. It should be noted, however, that only six independent samples contributed to the unpublished subgroup of effect sizes.

Third, we assessed whether the quality of studies influenced the strength of the effect. The moderating effect of quality was not significant, *Q*(1) = .31, *b* = −0.01, 95% CI = [−0.04, 0.03], *p* = .58, suggesting the included effect sizes were not biased by differences in methodological quality. Quality ratings for each study are presented in Table S1 in the Supplemental Material. Looking across included studies, marked strengths and weaknesses emerged—for example, nearly all studies clearly described hypotheses (98% of studies), tasks and measures (100%), and main findings (100%), but few reported power analyses (12%), withdrawals and dropouts (37%), or participant engagement with tasks (49%); see Supplemental Material Table S2 for descriptive statistics for all 18 items.

Finally, we ran a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014) to assess whether the *p*-value distribution for statistically significant (*p* < .05) effect

sizes in the meta-analysis aligned with the *p*-value distribution expected from a true effect. A *p* curve for a true effect should be right-skewed; it should contain more low (.01) than high (.04) significant *p* values (Simonsohn et al., 2014). The *p* curve for this meta-analysis (generated via the app at p-curve.com) was heavily right-skewed, indicating no evidence of publication bias or selective reporting of significant results in included studies (see Fig. S7 in the Supplemental Material). The *p*-curve analysis also provided an estimate of the statistical power of studies that yielded significant *p* values; for this meta-analysis, power was estimated to be 99%, 90% CI = [98%, 99%], indicating that these studies, on average, were well powered to detect true effects.

Discussion

Overgeneral memory has been well established in depression (van Vreeswijk & de Wilde, 2004; Williams et al., 2007), but whether overgenerality extends from past to future thinking has not, until now, been the focus of a comprehensive meta-analysis. By examining the currently available evidence from a range of sources—including six electronic databases—this meta-analysis provides the most complete account to date of the links between depression and the specificity of future thinking. We show how differences in methods and participant characteristics can have a profound impact on the strength of the effect.

We found that, on average, higher levels of depression were weakly correlated with reduced specificity of future thinking ($r = -.13$). Variation in future specificity could explain only 1.6% of the variation in levels of depression; however, the relationship was highly significant, indicating a small but reliable effect. Results also revealed substantial heterogeneity across the true effect sizes of included studies. To examine the source of this heterogeneity, we ran moderator analyses on several variables related to differences in samples and study designs. Although most effects were nonsignificant, we identified three key variables that had a significant effect on the strength of the relationship between depression and specificity: the emotional valence of future thinking, whether depression was measured categorically or dimensionally, and the sex of the participants. We discuss each significant moderator in turn before addressing the other variables.

Significant moderators

We predicted depression would be linked to reduced future specificity of all emotional valences but that the effect would be strongest for positive future thinking—a hypothesis based largely on prior empirical findings (e.g., Dickson & Bates, 2006; Stöber, 2000). Specificity was indeed most clearly reduced in depression for positive future thinking. The discrepancy in the effect across emotional valences was even larger than expected—for neutral future thinking, the effect was small and nonsignificant, and for negative future thinking, the effect disappeared entirely.

These findings have important implications because they run counter to the most influential model of overgenerality in depression: the CaR-FA-X model (Williams et al., 2007). This model describes how rumination, functional avoidance, and executive dysfunction disrupt the search and retrieval of specific events. Although originally applied to overgeneral memory, the same factors should also hinder the generation of specific future events (Dickson & Bates, 2005; Williams et al., 2007), especially in light of emerging neuroscientific evidence that episodic memories and future simulations are instantiations of the same underlying process (Addis, 2018). The CaR-FA-X model sets up specific predictions regarding valence, namely that functional avoidance should result in the truncation of negative future thinking to avoid unpleasant or painful fragments coming to mind (Williams et al., 2007). In contrast, other components in the model such as poor executive function should affect the specificity of future events irrespective of valence. Thus, by this account, future thinking of any valence is expected to be overgeneral in depression, but negative future thinking should be especially overgeneral.

Our finding that depressed and dysphoric people exhibited a significant reduction in specificity of positive but not negative events runs counter to the CaR-FA-X model. Rather, this pattern of results may be more parsimoniously explained—at least in part—by mood-congruence memory effects. Specifically, given that imagining specific future events relies on the retrieval of relevant details from memory (Schacter et al., 2012), it is plausible that the well-established reduction in the accessibility of positive memories with depressed/dysphoric mood (reviewed in Eich, 1995; Matt, Vázquez, & Campbell, 1992) could alter the ability to generate specific positive future events. One way to test this idea would be to examine overgenerality of future thinking in individuals with remitted depression; if mood congruence largely accounts for overgenerality, and these individuals are no longer in a negative-mood state, they should no longer exhibit the effect. Unfortunately, only two effect sizes in the meta-analysis were derived from samples exclusively of participants with remitted depression. Although these samples showed no overgenerality, we hesitate to draw inferences from so few cases. Future studies comparing the specificity of future

thinking in those with remitted versus current depression will help to resolve whether overgenerality can be better explained by trait or state (e.g., negative mood) factors.

Whatever the underlying mechanism, these findings have important clinical implications. Depressed and dysphoric individuals appear not to suffer a broad inability to imagine the future but rather struggle to imagine specific positive futures. It follows that the relatively intact ability to imagine in these individuals might be harnessed or directed toward positive future thinking in a way that could be beneficial. For example, a recent study showed that the generation of vivid, positive mental imagery for planned rewarding activities increased motivation for (and actual completion of) those activities relative to a nonimagery control condition (Renner, Murphy, Ji, Manly, & Holmes, 2019). Such effects might be particularly relevant for depressed and dysphoric individuals, who are less likely to plan and engage in rewarding activities, thus depriving themselves of potentially positive experiences and perpetuating low mood (Holmes, Blackwell, Burnett Heyes, Renner, & Raes, 2016). The findings of this meta-analysis suggest that the underlying capacity to vividly imagine the future is still present in depression and dysphoria—and perhaps can be repurposed more positively.

Of the other preregistered moderators, only the effect of categorical versus dimensional measures of depression was statistically significant. Studies with categorical designs (e.g., healthy control subjects vs. dysphoric subjects) showed a slightly stronger relationship between depression and future specificity than studies with dimensional designs (e.g., depression measured on a continuous scale such as the BDI-II). This is surprising statistically, as the dichotomization of a continuous predictor variable—which is what occurs in categorically designed studies—typically leads to reduced power and underestimation of the strength of relationships (Hunter & Schmidt, 1990; Maxwell & Delaney, 1993). A possible explanation is that this moderator was confounded by other variables. For instance, studies with categorical designs typically encompassed samples with a wider spectrum of depressive symptoms (e.g., spanning from healthy control subjects to patients with MDD) than those with dimensional designs (e.g., spanning from healthy control subjects to dysphoric participants). Samples spanning a wider spectrum of depression showed a slightly stronger (although not significantly different) effect than those spanning a narrower spectrum of depression, which may have contributed to the slightly larger effect in categorical than dimensional designs. This moderator analysis was the only one substantially affected by outliers, as removing the five most influential effect sizes (identified via Cook's D_i) shifted the result to nonsignificant. Thus, we hesitate to draw any strong conclusions about the moderating effect of categorical versus dimensional designs.

The third and final significant moderator variable was sex: The relationship between depression and specificity was weaker in samples with a higher proportion of females. Although the analysis was unplanned (exploratory), the effect was strong and significant and remained so after the removal of outliers. We are unaware of any previous suggestion that the effect of overgenerality may be stronger for depressed men than women. Some studies, however, have reported that women tend to generate more elaborate, detailed, and vivid autobiographical memories than men (Grysman & Hudson, 2013), and such a sex difference may extend to the imagination of future events (Wang, Hou, Tang, & Wiprovnick, 2011). If this is the case, perhaps women have more of a "buffer" against the effects of depressive symptomatology than men; that is, their ability to project richly and specifically into the future may remain intact for longer in depression. Interestingly, women are also more likely to engage in rumination than men (Nolen-Hoeksema & Jackson, 2001). If overgenerality in depression is driven partly by rumination, as the CaR-FA-X model suggests, specificity should be particularly reduced in women compared with men, and yet we found the opposite—possibly pointing to other causes of overgenerality as mentioned above. Our results also highlighted the relative lack of males in studies in this area (only 32% of the total 4,813 participants were men), suggesting future research should aim for more representative samples. We propose that further examination of sex differences in overgenerality is warranted: If men are especially affected in depression, men might also be especially amenable to future interventions that improve prospection.

Nonsignificant moderators

We now turn to the moderator variables that yielded nonsignificant effects. As predicted, the clinical status of participants did not have a significant effect on the link between depression and specificity—specificity was affected whether participants had subclinical depression (dysphoria) or a clinical diagnosis of MDD. As mentioned above, the effect was marginally stronger in samples that spanned healthy to MDD subjects versus healthy to dysphoric subjects, but the difference was nonsignificant. These findings are generally consistent with a dimensional model of depression, that is, that depression varies along a continuum of severity throughout the general population, with no sharp distinction in the symptoms of those with below- and

above-threshold depression (Ayuso-Mateos et al., 2010). If depression is truly continuous, one would also expect to see (small) correlations between specificity and depression within groups only with low to mild symptoms, or dysphoria, or MDD. Although we did not observe significant effects within these subgroups, correlations were similar in magnitude to the summary estimate. Considering this, and that these latter subgroups had only few cases, we cannot be sure whether the null findings reflect true null effects or simply insufficient power.

The two other moderator variables relating to participants—comorbid anxiety and age—also had nonsignificant effects. Few studies grouped or excluded participants on the basis of comorbid anxiety (each subgroup contained only six effect sizes), so little data were available for this analysis, and statistical power was low. It would be worthwhile to examine interactions between the moderators comorbid anxiety and emotional valence, as existing research suggests anxious individuals might show enhanced specificity for future negative events but not positive events (MacLeod, Tata, Kentish, & Jacobsen, 1997; MacLeod et al., 2005; Morina et al., 2011). Again, it was not feasible to investigate this question in the current study; interaction analyses would require more data.

As for age, we expected the effect to be stronger in older samples given future thinking in older depressed individuals may suffer compounding effects of normal aging and repeated depressive episodes (King et al., 2010; Schacter et al., 2013). That age was nonsignificant ran contrary to our prediction, but the lack of older samples in the meta-analysis may have reduced the likelihood to detect an effect. Surprisingly few studies examined overgenerality and depression in older adults (only five independent samples had a mean age greater than 50 years), suggesting this question could be a key avenue for future research.

The type of cue used to prompt participants to think about the future was also a nonsignificant moderator. Nonetheless, this analysis did reveal that, when considering effect sizes only from studies that used more supportive event cues, there was no evidence that specificity is reduced in depression. Event cues (e.g., "New Year's Eve" or "Christmas dinner") arguably provide a scaffold for accessing and generating more specific events than single-word cues (e.g., "party" or "dinner"; Addis et al., 2016). Whether the null result for event cues is due to insufficient data or whether a more supportive scaffold does indeed mitigate overgenerality in depression will need to be addressed in further studies.

That micro- versus macrolevel measures of specificity also had no significant moderating effect suggests overgenerality in depression is robust regardless of the level at which specificity is measured. In other words, specificity seems to be similarly reduced in depression in terms of the spatiotemporal specificity of events themselves as well as the episodic detail or vividness within those events. Similar findings emerged for whether specificity and depression were measured by the researcher or participant self-report and whether specificity pertained to future simulations, intentions, or plans. That is, reduced specificity was evident in depression regardless of differences in these methods or modes of future thinking—again suggesting this is a robust effect.

Are the results biased?

The overall effect we observed here for future thinking and depression ($r = -.13$) was far smaller than the moderate to large effect sizes reported in previous meta-analyses on overgeneral memory and depression. Williams et al. (2007) reported a mean Cohen's *d* of 1.12 across 12 studies, which translates to a biserial correlation of .61, and van Vreeswijk and de Wilde (2004), across 14 studies, reported Spearman's rho correlations of −.66 and .59 between depression and specific positive memories and depression and negative overgeneral memories, respectively. The latter meta-analysis included samples with comorbid psychiatric diagnoses (in addition to anxiety), however, which may have inflated effect sizes compared with samples with depression alone. In the recent meta-analysis on depression and future specificity by Hallford et al. (2018), the reported effect size was also large; across the seven included studies, Hedges's *g* was 0.79. It is difficult to explain the discrepancy in effect sizes, particularly as our meta-analysis included the seven studies examined by Hallford et al. (2018). In addition, one might have expected specificity in depression to be even more affected for future thinking than memory, as imagining novel scenes is more cognitively demanding than reconstructing the past (Addis et al., 2016). The discrepancies then raise the question: Are the results in the current meta-analysis affected by bias?

Across the four methods we used to assess publication and/or reporting bias, the results were mixed. On the one hand, the funnel plot (Fig. 4) revealed a strange pattern of a few very large and precise effect sizes—precision that would not be expected from the small samples in those studies, and we ensured these effect sizes did not arise from obvious misreporting or mistakes in data extraction. Trim-and-fill analysis suggested the meta-analysis may have missed additional studies with large effect sizes, meaning the true effect size might be larger. On the other hand, a moderator analysis showed that unpublished studies had weaker correlations on average than published studies, indicating

potential publication bias in the other direction; that is, weak or null effects might remain hidden in the "file drawer." And yet, clouding the picture further, differences in study quality were not related to the size of effects, and the *p* curve for significant results matched what would be expected from a true effect. So, overall, the mixed results make it difficult to draw conclusions about bias, although we find no particularly compelling evidence that the real effect size should be substantially different to the summary estimate. We have confidence in the current estimate because the search process was extensive, the number of included effect sizes and total *N* were relatively large, and the analyses included multiple measures of specificity that were corrected for dependency between outcomes. And, importantly, additional analyses were run without outliers to ensure conclusions were not affected by a few extreme studies.

Future specificity in depression: what else matters?

Although some moderators had a notable effect on the relationship between depression and specificity, none came close to explaining all or even most of the heterogeneity across studies. What else might account for some of this variability? As mentioned, complex interactions between moderators (such as comorbid anxiety and emotional valence) may have affected the relationship between future specificity and depression, but the small samples in some subgroups precluded our ability to run interaction analyses. Other factors not included in the current meta-analysis may have also had important moderating effects. Antidepressant medication, for example, could mitigate some of the effects of overgenerality in depression. Chronic antidepressant treatment has been found to have a neuroprotective effect on the hippocampus (Huang et al., 2013), a brain region critical for episodic memory and future thinking (Addis & Schacter, 2012) that tends to be reduced in volume in depression (Malykhin, Carter, Seres, & Coupland, 2010). Participants' history of depression may be another key moderator, as repeated depressive episodes over time could amplify the extent of overgenerality (King et al., 2010). It might also be valuable to examine the effect of participants' history of therapy, as common therapies such as cognitive-behavioral therapy often involve shifting clients' perceptions of the future (Seligman, Railton, Baumeister, & Sripada, 2013). Finally, given the multitude of tasks tapping into aspects of future specificity (26 different tasks were used across included studies), there were likely subtle differences between the tasks not captured by the current coding of methodological moderator variables.

To meta-analyze the effect of antidepressant medication, history of depression, or history of therapy would require either access to participant-level data or the categorizing of depressed participants into subgroups (such as medicated and nonmedicated), which was rarely done in the included studies. This situation is a clear example of the potential benefits of sharing (deidentified) participant-level data;[18] if this information had been collected and shared in some of the included studies, moderator analyses could be performed without the need to run additional studies. Moreover, running a "mega-analysis" (i.e., on participant- rather than study-level data) would reduce the chance of both false-positive and false-negative effects (Costafreda, 2009). With online tools such as the Open Science Framework now making it easier to share data, we hope it will be possible to conduct more powerful (and thus more illuminating) analyses on specificity in depression in the coming years.

Conclusion

Based on the currently available evidence derived from 46 studies, the results of this meta-analysis have demonstrated a small but robust association between higher levels of depression and the reduced specificity of future thinking. Although the magnitude of the effect was small, especially compared with previous meta-analyses on overgenerality in depression, we do not think it is necessarily trivial; even a small correlation may be relevant from a practical and clinical perspective when it relates to a variable as consequential as the severity of depression. Indeed, an intervention targeting future episodic specificity has already been shown to reduce negative affect, boost positive affect, and increase coping behaviors (Jing et al., 2016). This meta-analysis furthers our understanding of future thinking in depression and dysphoria by examining the people, and the conditions, for which overgenerality is most apparent; namely, specificity is most affected when attempting to imagine positive future events (and not at all affected when imagining negative events) and seems to be more severe in depressed men than women. We hope future interventions, guidelines, and policies can take advantage of these findings to highlight aspects of future thinking that are most in need of improvement. Finally, with our data and analysis scripts publicly available, we aim to make this meta-analysis transparent—and easily updated as new research comes to light.

future thinking in depression. D. Moreau provided statistical expertise. All the authors read, provided feedback on, and approved the final manuscript for submission. B. Gamble is the guarantor.

ORCID iD

Beau Gamble 🆔 https://orcid.org/0000-0001-8284-9371

Acknowledgments

Declaration of Conflicting Interests

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/174569161985178

Notes

1. Although there is a vast amount of research on the different types of future-oriented cognitions (e.g., Bandura, 1986; Locke & Latham, 1990; Mischel, 1973; Oettingen & Mayer, 2002; Taylor & Schneider, 1989), this taxonomy (Szpunar, Spreng, & Schacter, 2014) is intended to encompass the majority of these cognitions, and indeed we found that all studies included in the meta-analysis could be categorized into one of the four modes of prospection.

2. The Oxford English Dictionary defines "vivid" (n.d.) as "producing powerful feelings or strong, clear images in the mind."

3. Although this distinction is typically made for future events, we also coded the emotional valence of other types of future simulations, such as goals and plans—if the distinction was clear. For example, approach goals were classified as *positive*, as they reflect movement toward a desired positive state, and avoidance goals as *negative*, as they reflect movement away from an unwanted negative state (Elliot & Sheldon, 1997).

4. Likewise, in addition to events, we applied the macro- versus microlevel distinction to measures of specificity of other types of future simulations, such as goals and plans (Dickson & MacLeod, 2004; Emmons, 1992).

5. We noted that differences in cue type may be confounded by other differences in scoring method (e.g., studies using single-word cues typically use macrolevel measures of specificity)—which could make these effects difficult to tease apart.

6. The PICOS categories *interventions* and *comparisons* were excluded because they cannot be applied to meta-analyses of observational studies.

7. For example, *specific* is synonymous with *definite*, which is synonymous with *concrete*; we hence took measures of the concreteness of future thinking as capturing some aspect of specificity.

8. In the Future Thinking Task participants are asked to orally name as many future events they are looking forward to, and

not looking forward to, as possible, with the typical outcome measure being the total number of named events.

9. The syntax used to search each database can be downloaded from https://osf.io/dk6ws.

10. Full reports of two potentially relevant articles could not be obtained, and the authors did not respond to e-mails.

11. For one effect size (the positive valence condition in Dickson & Bates, 2006) the conversion yielded a nonsensical value of $r = -1.05$. Rather than excluding this effect size we converted Cohen's d to a point-biserial correlation, which assumes the groups were categorically different in some way, yielding a value of $r = -.83$.

12. Most studies in the meta-analysis either did not measure anxiety or did not group participants according to anxiety levels.

13. Random-effects meta-analysis (unlike fixed-effects meta-analysis) do not assume the true effect size is the same across included studies, so it is the preferred option when heterogeneity is present (Riley, Higgins, & Deeks, 2011).

14. Restricted maximum likelihood is the standard option for this analysis and thought to be relatively unbiased (Viechtbauer, 2010).

15. It was planned that B. Gamble and D. Moreau would assess quality ratings independently, but for practical purposes, the authors assessed quality in consultation, and the values reported are consensus scores.

16. k denotes the number of effect sizes in each subgroup.

17. We are unaware of any studies on the "specificity of prediction"; perhaps this would entail ratings not only on the probability of future events but also on the participant's confidence around those predictions.

18. Of course, such data sharing is possible only when participants have provided informed consent and when data sharing is permissible by a country's relevant privacy laws.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, *48*(2–3), 64–88. doi:10.1080/03036758.2018.1439071

*Addis, D. R., Hach, S., & Tippett, L. J. (2016). Do strategic processes contribute to the specificity of future simulation in depression? *British Journal of Clinical Psychology*, *55*, 167–186. doi:10.1111/bjc.12103

Addis, D. R., & Schacter, D. L. (2012). The hippocampus and imagining the future: Where do we stand? *Frontiers in Human Neuroscience*, *5*, Article 173. doi:10.3389/fnhum .2011.00173

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science*, *19*, 33–41. doi:10.1111/j.1467-9280 .2008.02043.x

Anderson, R. J., Boland, J., & Garner, S. R. (2016). Overgeneral past and future thinking in dysphoria: The role of emotional cues and cueing methodology. *Memory*, *24*, 708–719. doi:10.1080/09658211.2015.1046134

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, *5*, 533–539. doi:10.1016/S1364-6613(00)01804-0

Ayuso-Mateos, J. L., Nuevo, R., Verdes, E., Naidoo, N., & Chatterji, S. (2010). From depressive symptoms to depressive disorders: The relevance of thresholds. *British Journal of Psychiatry*, *196*, 365–371. doi:10.1192/bjp.bp.109.071191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Beck, A. (1976). *Cognitive therapy and the emotional disorders*. New York, NY: International Universities Press.

Beck, A., Ward, C., Mendelson, M., & Mock, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Becker, G. (1986). Correcting the point-biserial correlation for attenuation owing to unequal sample size. *Journal of Experimental Education*, *55*, 5–8. doi:10.1080/00220973.1986.10806427

Benoit, R. G., & Schacter, D. L. (2015). Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia*, *75*, 450–457. doi:10.1016/j.neuropsychologia.2015.06.034

*Boelen, P. A., Huntjens, R. J. C., & van den Hout, M. A. (2014). Concurrent and prospective associations of habitual overgeneral memory and prospection with symptoms of depression, general anxiety, obsessive compulsiveness, and post-traumatic stress. *Memory*, *22*, 747–758. doi:10.1080/09658211.2013.824985

Borkovec, T. D., & Ray, W. J. (1998). Worry: A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research*, *22*, 561–576.

Brittlebank, A., Scott, J., Williams, J., & Ferrier, I. (1993). Autobiographical memory in depression: State or trait marker? *The British Journal of Psychiatry*, *162*, 118–121. doi:10.1192/bjp.162.1.118

Brown, A. D., Root, J. C., Romano, T. A., Chang, L. J., Bryant, R. A., & Hirst, W. (2013). Overgeneralized autobiographical memory and future thinking in combat veterans with posttraumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 129–134. doi:10.1016/j.jbtep.2011.11.004

Brown, G. P., MacLeod, A. K., Tata, P., & Goddard, L. (2002). Worry and the simulation of future outcomes. *Anxiety, Stress & Coping*, *15*, 1–17. doi:10.1080/10615800290007254

Costafreda, S. (2009). Pooling fMRI data: Meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, *3*, Article 33. doi:10.3389/neuro.11.033.2009

Cropley, M., & MacLeod, A. K. (2003). Dysphoria, attributional reasoning and future event probability. *Clinical Psychology and Psychotherapy*, *10*, 220–227. doi:10.1002/cpp.360

D'Argembeau, A., Ortoleva, C., Jumentier, S., & Van der Linden, M. (2010). Component processes underlying future thinking. *Memory & Cognition*, *38*, 809–819. doi:10.3758/MC.38.6.809

Dickson, J. M., & Bates, G. W. (2005). Influence of repression on autobiographical memories and expectations of the future. *Australian Journal of Psychology*, *57*, 20–27. doi:10.1080/00049530412331283444

*Dickson, J. M., & Bates, G. W. (2006). Autobiographical memories and views of the future: In relation to dysphoria. *International Journal of Psychology*, *41*, 107–116. doi:10.1080/00207590500188025

*Dickson, J. M., & MacLeod, A. (2004). Approach and avoidance goals and plans: Their relationship to anxiety and depression. *Cognitive Therapy and Research*, *28*, 415–432. doi:10.1080/02699930341000013

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, *52*, 377–384. doi:10.1136/jech.52.6.377

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98. doi:10.1080/01621459.2000.10473905

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

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science*, *6*, 67–75.

Elliot, A. J., & Sheldon, K. M. (1997). Avoidance achievement motivation: A personal goals analysis. *Journal of Personality and Social Psychology*, *73*, 171–185. doi:10.1037/0022-3514.73.1.171

*Emmons, R. A. (1992). Abstract versus concrete goals: Personal striving level, physical illness, and psychological well-being. *Journal of Personality and Social Psychology*, *62*, 292–300.

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, *58*, 33–48. doi:10.1016/j.cpr.2017.09.005

Farrin, L., Hull, L., Unwin, C., Wykes, T., & David, A. (2003). Effects of depressed mood on objective and subjective measures of attention. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *15*, 98–104. doi:10.1176/jnp.15.1.98

First, M., Spitzer, R., Williams, J., & Gibbon, M. (1995). *Structured Clinical Interview for DSM-IV–Non-Patient Edition (SCID-NP, Version 1.0)*. Washington, DC: American Psychiatric Association.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., . . . Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*, *64*, 1187–1197. doi:10.1016/j.jclinepi.2010.08.010

Grysman, A., & Hudson, J. A. (2013). Gender differences in autobiographical memory: Developmental and methodological considerations. *Developmental Review*, *33*, 239–272. doi:10.1016/j.dr.2013.07.004

*Hach, S., Tippett, L. J., & Addis, D. R. (2016). *A closer examination of episodic specificity in depression*. 6th International Conference on Memory, Budapest, Hungary. Retrieved from https://hdl.handle.net/10652/3735

*Hadley, S. A., & MacLeod, A. K. (2010). Conditional goal-setting, personal goals and hopelessness about the future. *Cognition & Emotion*, *24*, 1191–1198. doi:10.1080/02699930903122521

Hallford, D. J., Austin, D. W., Takano, K., & Raes, F. (2018). Psychopathology and episodic future thinking: A systematic review and meta-analysis of specificity and episodic detail. *Behaviour Research and Therapy*, *102*, 42–51. doi:10.1016/j.brat.2018.01.003

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Holmes, E. A., Blackwell, S. E., Burnett Heyes, S., Renner, F., & Raes, F. (2016). Mental imagery in depression: Phenomenology, potential mechanisms, and treatment implications. *Annual Review of Clinical Psychology*, *12*, 249–280. doi:10.1146/annurev-clinpsy-021815-092925

*Holmes, E. A., Lang, T. J., Moulds, M. L., & Steele, A. M. (2008). Prospective and positive mental imagery deficits in dysphoria. *Behaviour Research and Therapy*, *46*, 976–981. doi:10.1016/j.brat.2008.04.009

Huang, Y., Coupland, N. J., Lebel, R. M., Carter, R., Seres, P., Wilman, A. H., & Malykhin, N. V. (2013). Structural changes in hippocampal subfields in major depressive disorder: A high-field magnetic resonance imaging study. *Biological Psychiatry*, *74*, 62–68. doi:10.1016/j.biopsych.2013.01.005

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*, 334–349. doi:10.1037/0021-9010.75.3.334

Jing, H. G., Madore, K. P., & Schacter, D. L. (2016). Worrying about the future: An episodic specificity induction impacts problem solving, reappraisal, and well-being. *Journal of Experimental Psychology: General*, *145*, 402–418. doi:10.1037/xge0000142

*King, M. J., MacDougall, A. G., Ferris, S., Herdman, K. A., & McKinnon, M. C. (2011). Episodic simulation of future events is impaired in patients with major depressive disorder. *Psychiatry Research*, *187*, 465–467. doi:10.1016/j.psychres.2011.02.002

King, M. J., MacDougall, A. G., Ferris, S. M., Levine, B., MacQueen, G. M., & McKinnon, M. C. (2010). A review of factors that moderate autobiographical memory performance in patients with major depressive disorder. *Journal of Clinical and Experimental Neuropsychology*, *32*, 1122–1144. doi:10.1080/13803391003781874

Kleim, B., Graham, B., Fihosy, S., Stott, R., & Ehlers, A. (2014). Reduced specificity in episodic future thinking in post-traumatic stress disorder. *Clinical Psychological Science*, *2*, 165–173. doi:10.1177/2167702613495199

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Mackinger, H. F., Pachinger, M. M., Leibetseder, M. M., & Fartacek, R. R. (2000). Autobiographical memories in women remitted from major depression. *Journal of Abnormal Psychology*, *109*, 331–334. doi:10.1037//0021-843X.109.2.331

*MacLeod, A. K., & Cropley, M. (1995). Depressive future-thinking: The role of valence and specificity. *Cognitive Therapy and Research*, *19*, 35–50. doi:10.1007/BF02229675

MacLeod, A. K., Tata, P., Kentish, J., & Jacobsen, H. (1997). Retrospective and prospective cognitions in anxiety and depression. *Cognition & Emotion*, *11*, 467–479. doi:10.1080/026999397379881

MacLeod, A. K., Tata, P., Tyrer, P., Schmidt, U., Davidson, K., & Thompson, S. (2005). Hopelessness and positive and negative future thinking in parasuicide. *The British Journal of Clinical Psychology*, *44*, 495–504. doi:10.1348/014466505X35704

Malykhin, N. V., Carter, R., Seres, P., & Coupland, N. J. (2010). Structural changes in the hippocampus in major depressive disorder: Contributions of disease and treatment. *Journal of Psychiatry and Neuroscience*, *35*, 337–343. doi:10.1503/jpn.100002

Martin, V. C. (2013). *Memory for the future: The encoding and phenomenology of episodic simulations* (Unpublished doctoral dissertation). The University of Auckland, Auckland, New Zealand.

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

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*, 181–190. doi:10.1037/0033-2909.113.1.181

McNally, R. J., Lasko, N. B., Macklin, M. L., Pitman, R. K., & McNally, R. J. (1995). Autobiographical memory and combat PTSD. *Behavioral Research and Therapy*, *33*, 619–630. doi:10.1016/0005-7967(95)00007-K

McNally, R. J., Litz, B. T., Prassas, A., Shin, L. M., & Weathers, F. W. (1994). Emotional priming of autobiographical memory in post-traumatic stress disorder. *Cognition & Emotion*, *8*, 351–367. doi:10.1080/02699939408408946

Miloyan, B., Bulley, A., & Suddendorf, T. (2016). Episodic foresight and anxiety: Proximate and ultimate perspectives. *British Journal of Clinical Psychology*, *55*, 4–22. doi:10.1111/bjc.12080

Mischel, W. (1973). Toward a cognitive social learning. *Psychological Review*, *80*, 252–283. doi:10.1037/h0035002

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), Article 1. doi:10.1186/2046-4053-4-1

Moreau, D., & Corballis, M. C. (2019). When averaging goes wrong: The case for mixture model estimation in psychological science. *Journal of Experimental Psychology: General*. Advance online publication. doi:10.1037/xge0000504

*Morina, N., Deeprose, C., Pusowski, C., Schmid, M., & Holmes, E. A. (2011). Prospective mental imagery in patients with major depressive disorder or anxiety disorders. *Journal of Anxiety Disorders*, *25*, 1032–1037. doi:10.1016/j.janxdis.2011.06.012

Nelson, K. L., Moskovitz, D. J., & Steiner, H. (2008). Narration and vividness as measures of event-specificity in autobiographical memory. *Discourse Processes*, *45*, 195–209. doi:10.1080/01638530701792891

Nolen-Hoeksema, S., & Jackson, B. (2001). Mediators of the gender difference in rumination. *Psychology of Women Quarterly*, *25*, 37–47. doi:10.1111/1471-6402.00005

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology*, *83*, 1198–1212. doi:10.1037/0022-3514.83.5.1198

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Journal of Applied Psychological Measurement*, *1*, 385–401.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Renner, F., Murphy, F. C., Ji, J. L., Manly, T., & Holmes, E. A. (2019). Mental imagery as a "motivational amplifier" to promote activities. *Behaviour Research and Therapy*, *114*, 51–59. doi:10.1016/j.brat.2019.02.002

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *British Medical Journal*, *342*, 2–5. doi:10.1136/bmj.d549

*Robinaugh, D. J., Lubin, R. E., Babic, L., & McNally, R. J. (2013). Are habitual overgeneral recollection and prospection maladaptive? *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 227–230. doi:10.1016/j.jbtep.2012.11.002

Roepke, A. M., & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology*, *55*, 23–48. doi:10.1111/bjc.12087

Sarkohi, A. (2011). Links between future thinking and autobiographical memory specificity in major depression. *Psychology*, *2*, 260–264. doi:10.4236/psych.2011.23041

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*, 657–661. doi:10.1080/08995600802554748

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*, 677–694. doi:10.1016/j.neuron.2012.11.001

Schacter, D. L., Gaesser, B., & Addis, D. R. (2013). Remembering the past and imagining the future in the elderly. *Gerontology*, *117*, 247–251. doi:10.1159/000342198

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). Small-study effects in meta-analysis. *In Meta-analysis with R* (pp. 107–142). Cham, Switzerland: Springer.

Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8*, 119–141. doi:10.1177/1745691612474317

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534–547. doi:10.1037/a0033242

*Stöber, J. (2000). Prospective cognitions in anxiety and depression: Replication and methodological extension. *Cognition & Emotion*, *14*, 725–729. doi:10.1080/02699930050117693

Street, H. (2002). Exploring relationships between goal setting, goal pursuit and depression: A review. *Australian Psychologist*, *37*, 95–103. doi:10.1080/00050060210001706736

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences, USA*, *111*, 18414–18421. doi:10.1073/pnas.1417144111

Taylor, S. E., Pham, L. B., Rivkin, I. D., & Armor, D. A. (1998). Harnessing the imagination: Mental simulation, self-regulation, and coping. *The American Psychologist*, *53*, 429–439. doi:10.1037/0003-066X.53.4.429

Taylor, S. E., & Schneider, S. K. (1989). Coping and the simulation of events. *Social Cognition*, *7*, 174–194. doi:10.1521/soco.1989.7.2.174

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York, NY: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12. doi:10.1037/h0080017

van Vreeswijk, M. F., & de Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy*, *42*, 731–743. doi:10.1016/S0005-7967(03)00194-3

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

Vivid. (n.d.). In *Lexico*. Retrieved from https://www.lexico.com/en/definition/vivid

Wang, Q., Hou, Y., Tang, H., & Wiprovnick, A. (2011). Travelling backwards and forwards in time: Culture and gender in the episodic specificity of past and future events. *Memory*, *19*, 103–109. doi:10.1080/09658211.2010.537279

Williams, J., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*, 122–148. doi:10.1037/0033-2909.133.1.122

Williams, J., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology*, *95*, 144–149. doi:10.1037/0021-843X.95.2.144

*Williams, J., Ellis, N. C., Tyers, C., Healy, H., Rose, G., & MacLeod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Memory & Cognition*, *24*, 116–125. doi:10.3758/BF03197278

Williams, J., & Scott, J. (1988). Autobiographical memory in depression. *Psychological Medicine*, *18*, 689–695. doi:10.1017/S0033291700008370

Wu, J. Q., Szpunar, K. K., Godovich, S. A., Schacter, D. L., & Hofmann, S. G. (2015). Episodic future thinking in generalized anxiety disorder. *Journal of Anxiety Disorders*, *36*, 1–8. doi:10.1016/j.janxdis.2015.09.005