A Meta-Analytic Review of the Relationship Between Explicit Memory Bias and Depression: Depression Features an Explicit Memory Bias That Persists Beyond a Depressive Episode

Jonas Everaert1, 2, Janna N. Vrijen3, 4, Renée Martin-Willett5, Livia van de Kraats3, and Jutta Joormann6

1 Department of Medical and Clinical Psychology, Tilburg University
2 Research Group of Quantitative Psychology and Individual Differences, KU Leuven
3 Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center
4 Depression Expertise Center, Pro Persona Mental Health Care, Nijmegen, Gelderland, The Netherlands
5 Department of Psychology and Neuroscience, University of Colorado Boulder
6 Department of Psychology, Yale University

Emotional bias in explicit memory is theorized to play a prominent role in the etiology, maintenance, and recurrence of depression. Even though this cognitive bias is regarded as one of the most robust phenomena in depression, its magnitude and boundary conditions in depression are currently unknown. This review presents two three-level meta-analyses to estimate the overall effect size and identify moderators of explicit memory bias in depression. Meta-analysis I (153 studies, 686 contrasts) revealed a small overall effect size for naturalistic explicit memory bias in depression, $g = 0.241$, 95% CI [0.179, 0.304]. The magnitude of the overall effect was moderated by emotional valence of stimuli, operational definition of memory bias, depth of processing during encoding, explicit memory task, and the (non-)verbal nature of stimuli. Equivalent effect sizes were found for minors and adults as well as for clinical and subclinical depression. Remarkably, a nonsignificant effect size emerged for remitted depression. Following up on the latter finding, Meta-analysis II (21 studies, 80 contrasts) examined explicit memory bias in remitted depression under naturalistic conditions and under mood/stress induction. Results yielded a nonsignificant overall effect size, $g = 0.131$, 95% CI [0.045, 0.307], but a significant effect size for study conditions with mood or stress induction, $g = 0.273$, 95% CI [0.004, 0.542]. Both meta-analyses indicated high levels of heterogeneity, even after accounting for variation explained by sample and study characteristics. The findings are consistent with the view that depression is characterized by an explicit memory bias that may persist beyond a depressive episode. These findings have implications for cognitive theories of vulnerability to depression as well as clinical interventions.

Public Significance Statement
This meta-analytic review suggests that individuals with current depressive symptoms exhibit a bias in how they remember emotional information. This bias occurs particularly when individuals with depression try to remember positive information. Individuals who recovered from depression may show a memory bias only when experiencing a negative mood or feeling stressed. For persons who currently suffer from depressive symptoms and for persons who recovered but feel down or stressed, interventions may focus on improving difficulties in retrieving positive memories such as by repeated training of positive memory.

Keywords: memory bias, explicit memory, cognitive bias, depression, meta-analysis

Supplemental materials: https://doi.org/10.1037/bul0000367.supp

Depression is a highly prevalent mental health condition that causes deep personal suffering and has high societal costs (Kessler & Bromet, 2013). Globally, depression is the leading cause of disability, affecting more than 300 million people (World Health Organization, 2017). Despite a range of therapeutic interventions, relapse, recurrence, and nonresponse rates remain high (Munder et al., 2019; Vittengl et al., 2007). These facts highlight the importance of identifying etiological and maintenance factors for depression toward advancing prevention and treatment.

Cognitive models have assigned a prominent role to emotional bias in memory in the etiology, maintenance, and relapse of depression (A. T. Beck & Haigh, 2014; Bower, 1981; Clark et al., 1999; Hertel, 2004a; Ingram, 1984; J. M. G. Williams et al., 1988, 1997).
These models share the premise that depression is characterized by enhanced retrieval of negative emotional content and diminished retrieval of positive emotional content from long-term memory (LTM). This mood-congruency effect in LTM retrieval is theorized to occur in both explicit memory (i.e., more effortful and controlled storage of consciously accessible knowledge) and implicit memory (i.e., storing nonconsciously accessible knowledge expressed through behavior or performance; Gaddy & Ingram, 2014). The LTM bias in depression is expected to be particularly strong for explicit memory (J. M. G. Williams et al., 1997). Indeed, explicit LTM bias is regarded as a proximal cognitive vulnerability marker of depression and not merely as a mood-dependent correlate (A. T. Beck & Haigh, 2014; Clark et al., 1999; Connolly et al., 2016; Goldstein et al., 2015; Hayden et al., 2013; Johnson et al., 2007). Therefore, explicit LTM bias is an important target in emerging cognitive training interventions (Becker et al., 2015; Visser et al., 2020) and therapies (Edwards, 2007; Young et al., 2003).

Prior Research and Reviews

The prediction that depression is linked to a mood-congruent bias in explicit LTM has generated a wealth of empirical research. This work has employed a variety of explicit memory paradigms in different populations to document the nature and boundary conditions of explicit LTM bias in depression (for descriptive reviews, see Bogie et al., 2019; LeMoult & Gotlib, 2019; Wisco, 2009). The first, and until now, only meta-analysis on this topic was conducted in 1992 and integrated data from 14 early studies. The authors reported a small but significant effect size, suggesting that individuals with clinical depression are more likely to recall more negative than positive stimuli (Matt et al., 1992). However, this meta-analysis did not find evidence for a mood-congruent LTM bias in individuals with subclinical forms of depression, as they equally recalled negative and positive information. Taken together, these observations suggest that a negative bias or lack of positive bias in explicit LTM is particularly characteristic of clinical forms of depression.

Since the publication of this influential meta-analysis, contradictory findings have emerged in this literature. In contrast to Matt and colleagues’ observations that there may not be a memory bias in subclinical forms of depression (Matt et al., 1992), later research has reported evidence for biased recall of emotional information in samples of undiagnosed individuals reporting elevated depressive symptoms (Bianchi et al., 2020; Ingram, 1984; Koster et al., 2010; Moilanen, 1993; but also see Cooper & Wade, 2015; Yovel & Mineka, 2004). In addition, a recent review found that only a small subset of studies observed an LTM bias in clinical depression (Bogie et al., 2019). Mixed findings have also emerged from studies examining LTM bias in individuals who remitted from depression. Some studies observed a naturalistic memory bias in remitted depression (Bogie et al., 2020; Gethin et al., 2017; Hedlund & Rude, 1995; Romero et al., 2014) or a bias after providing a negative mood induction (Fritzsch et al., 2010; Timbremont & Braet, 2004; Vrijens et al., 2014). In contrast, several other studies did not find support for a mood-congruent memory bias in remitted depression (Arnold et al., 2011; Bradley & Mathews, 1988; Ruhe et al., 2019; Sears et al., 2011; Shestyk & Deldin, 2010; van Wingen et al., 2010), even when using mood induction (Ramel et al., 2007; Timbremont et al., 2008).

Inconsistencies within and between studies on subclinical, clinical, and remitted depression have emerged when operationalizing LTM bias as either a within-subject or a between-subject bias. Studies have reported that individuals with depression differ from nondepressed individuals in memory for negative stimuli (i.e., a between-subject bias), while depressed individuals equally recalled negative and neutral stimuli (i.e., no within-subject bias; Clifford & Hemsley, 1987; Ellis et al., 2014; Hamilton & Gotlib, 2008; Kuiper et al., 1985). Conversely, studies have found that individuals with depression recalled more negative than neutral information (i.e., a within-subject bias), but did not consistently differ from nondepressed individuals (i.e., no between-subject bias; Bentall & Kaney, 1996; Bradley & Mathews, 1983; Danion et al., 1995; Denny & Hunt, 1992; Gotlib & McCann, 1984). Though it remains unclear whether depression is characterized by a within-subject bias, a between-subjects bias, or both, studies have still interpreted their respective findings as support for a mood-congruent bias. This limitation renders conclusions about the strength of the overall effect size of explicit LTM bias in depression difficult.

Mixed findings have also emerged from research examining boundary conditions of explicit LTM bias by experimentally manipulating procedural features of laboratory tasks. These laboratory tasks typically require participants to encode a list of words or pictures with negative, neutral, or positive valence. During encoding, participants are instructed to study the presented stimuli, read the words aloud or watch the pictures, or imagine a scene involving themselves and the stimuli. Following a retention interval, participants are asked to remember the encoded words during a recall or recognition task. Varying the nature of this explicit memory test, studies did not find consistent evidence for explicit LTM bias across different tests, even within studies using multiple memory tests (Cerny et al., 2019; Kurtz & Morey, 1999; Liu et al., 2012; Neshat-Doost et al., 1998; Z. Zupan et al., 2017; but also see Auerbach et al., 2015; Rude et al., 1988).

While studies using free-recall tasks yielded relatively consistent evidence for explicit memory bias (Deldin et al., 2009; Dozois & Dobson 2001a; Kwiatkowski & Parkinson, 1994; Orchard & Reynolds, 2018; but also see Ellwart et al., 2003; Hsu et al., 2010; Reid et al., 2006), findings from studies using other memory tasks such as recognition have been less consistent (Deijen et al., 1993; Ellis et al., 2011; Jermann et al., 2009; Leal et al., 2014; Olsen et al., 2015; Ramponi et al., 2010; Sears et al., 2010). Additionally, investigators have manipulated the depth of encoding to examine the impact on memory bias by providing conceptual versus perceptual encoding instructions. Though some studies found no difference (Baños et al., 2001; Derry & Kuiper, 1981; Hammen & Zupan, 1984), others reported that conceptual but not perceptual encoding resulted in an LTM bias in depression (Ruiz-Caballero & González, 1997; B. A. Zupan et al., 1987). Moreover, prior work has suggested that LTM bias mainly occurs following self-referential processing (Wisco, 2009), but it remains unclear whether a memory bias is only elicited by self-referential conceptual processing of emotional information (Smith et al., 2018; Z. Zupan et al., 2017). It thus remains difficult to draw firm conclusions about the factors that modulate the strength of explicit LTM bias in depression.

Though explicit LTM bias in depression is often considered one of the strongest and most consistent findings in the literature on cognitive biases in depression (Gotlib & Joormann, 2010; LeMoult & Gotlib, 2019; Mathews & MacLeod, 2005), the inconsistencies that
have emerged challenge the understanding of this purported cognitive vulnerability factor in depression. The field lacks a comprehensive meta-analysis that synthesizes the large body of research on explicit memory bias in depression that has emerged in the past four decades. The current review aimed to fill this gap by estimating the magnitude of explicit memory bias and investigating potential moderators that could explain variability in research findings.

Cognitive Models of Memory Bias in Depression

To contextualize the variables of interest to this meta-analysis, this section briefly outlines several influential cognitive models of depression that have guided research on explicit memory bias. The models reviewed below propose several factors that modulate the nature and boundary conditions of the emotional bias in explicit memory.

One of the most influential cognitive models is Beck’s cognitive theory of depression (A. T. Beck & Haigh, 2014; Clark et al., 1999). Beck’s model asserts that individuals vulnerable to depression hold latent negative schemas that are stable cognitive structures or memory representations containing dysfunctional beliefs about the self (e.g., on themes of personal loss, failure, and deprivation). These negative schemas are developed in response to adverse or traumatic experiences during childhood. When activated by stressful events, these negative schemas will guide how emotional information in the environment is processed. Activated negative schemas are expected to bias information processing in favor of emotional information that is congruent with the negative schemas. Specifically, individuals with vulnerability to depression will selectively attend to negative cues in their environment, interpret ambiguous information as more negative, and recall self-relevant negative memories. These negative biases of attention, interpretation, and memory produce a stream of more negative and fewer positive cognitions about the self, others, and the world that engender depressive symptoms. It is proposed that the magnitude of these negative biases is a linear function of depressive symptoms. Thus, as depressive symptoms develop, a cognitive shift is expected to occur from a positivity bias to facilitated processing of negative information in clinically depressed individuals.

Drawing on Bower’s network theory (Bower, 1981), Ingram proposed an information-processing analysis attributing a central role to memory bias in self-perpetuating cycles of negative thought in depression (Ingram, 1984). The theory postulates that appraisals of life events activate depressive networks stored in memory. These memory networks consist of interconnected nodes containing sets of negative cognitions. Once a memory node is activated, the activation spreads to adjacent nodes within the memory network as well as connected memory networks. This selective spreading of activation to connected nodes and networks results in enhanced elaboration on negative memory representations. The process of recycling negative cognitions through various negative memory networks results in a deeper encoding of the elaborated material into the depressive memory networks. This encoding, in turn, increases the chances that this negative memory content becomes activated in the future. Ingram’s theory asserts that this biased elaboration–memory interaction is a vulnerability factor for depression that endures beyond the depressive episode.

Another influential cognitive model of depression is Williams et al.’s enhanced elaboration account (J. M. G. Williams et al., 1988, 1997). This account distinguishes between two stages of information processing: priming and elaboration. Priming refers to the early automatic internal representation of information, enhancing its accessibility. Elaboration refers to a strategic process strengthening the relation between such internal negative representations, resulting in more optimal encoding and later retrieval of negative information. Williams et al. hypothesized that depression is characterized by negative biases in elaboration. Two mechanisms are theorized to underlie this negativity bias, namely the affective decision mechanism (ADM) and the resource allocation mechanism (RAM). When the valence of incoming information is considered negative, as assessed by the ADM, more attention resources are allocated to negative material leading to enhanced elaboration (RAM). These depression-related elaborations are encoded into memory, enhancing later memory for depression-related material. According to the reformulated model (J. M. G. Williams et al., 1997), depressed individuals engage in strategic or biased elaboration of negative material during memory retrieval. This enhanced elaboration results in improved memory for similar information, and, in addition, such elaborations can serve as mnemonic cues at later points in time.

A final cognitive framework that has been influential in research on cognition and memory processes in depression is Hertel’s habits of thought framework (Hertel, 2004a, 2004b; Hertel & Brozovich, 2010). Hertel conceptualizes memory bias as a product of cognitive habits. Habits of thought refer to a ruminative processing style that is initiated automatically without conscious awareness or effort (Hertel, 2004b). This enhanced negative thinking results in repeated retrieval of negative material that strengthens negative memory bias as a habit in remembering. Interestingly, Hertel’s framework elaborates on when cognitive habits occur. The framework proposes that habits occur particularly in situations where stimuli and task dimensions do not inherently constrain and guide attention (Hertel, 1998, 2004b; Watkins & Nolen-Hoeksema, 2014). By contrast, habits of thought are expected to occur less likely in situations or experimental conditions that are constrained and force participants to sustain attention on task stimuli. When manifested, depressive habits of thought exacerbate sad mood and predict depressive episodes.

Moderating Variables Derived From Theory and Methodology

Empirical studies on explicit memory bias in depression differ widely in their focus on certain memory bias characteristics, sample characteristics, and memory task characteristics. Examining such characteristics as moderating variables may help to address the following key questions about explicit memory bias in depression: (a) How does explicit memory bias occur in depression? (b) Who exhibits an explicit memory bias? And (c), what are boundary conditions of explicit memory bias in depression?

How Does Explicit Memory Bias Occur in Depression?

Operational Definition

Research has adopted two operational definitions of memory bias, namely between-subject and within-subject bias. Between-subject
bias refers to differences in emotional LTM retrieval between individuals with depression and individuals without depression. Within-subject bias refers to the differences between the processing of emotional (i.e., negative or positive) and neutral stimuli in individuals with depression. Evidence for either a between-subject or a within-subject bias has been interpreted as support for the notion of explicit LTM bias in depression. However, there have been inconsistencies within and between studies in support of a between-subject or within-subject bias. The presence of a within-subject bias in the absence of a between-subject bias suggests that the bias may not be characteristic for individuals with depression. Conversely, the presence of a between-subject bias in the absence of a within-subject bias indicates that individuals with depression show no memory bias, but that individuals without depression remember emotional material in a positively biased manner. To date, a formal quantitative analysis of within-subject and between-subject bias in LTM in depression is missing but seems of crucial importance to accurately characterize the nature of memory bias in depression.

Content Specificity of Explicit Memory Bias

Depression is theorized to have a distinct cognitive profile that is evident in the content of negative cognitions and information-processing biases (Clark et al., 1999; J. M. G. Williams et al., 1997), reflecting themes of personal loss, failure, and deprivation (R. Beck & Perkins, 2001). At the level of cognitive processing, depression is theorized to be characterized by enhanced processing of negative self-referent information and minimization of positive self-referent material (A. T. Beck & Haigh, 2014; Clark et al., 1999). This cognitive profile of depression is thought to be distinct from anxiety, a disorder that features an excessive focus on physical or psychological threat or danger. According to the cognitive content-specificity hypothesis, individuals with depression exhibit a memory bias for depression-relevant information but not for threat-related information. However, previous work has reported a memory bias for some threat-related material in depression (Gotlib et al., 2004; but see Lim & Kim, 2005; Rinck & Becker, 2005; Tarsia et al., 2003), thus clarifying the specificity of explicit LTM bias in depression is important to developing cognitive interventions targeting information-processing biases.

Who Exhibits an Explicit Memory Bias?

Gender

Research indicates that women are nearly twice as likely to suffer from depression compared to men (Salk et al., 2017). Cognitive factors such as negative cognitive schemas and rumination may be more prevalent among women (Bone et al., 2020; Kuehner, 2017), and there may be a stronger relationship between self-referential negative memories and depressive symptoms among adolescent females compared to males (Bone et al., 2021). Indeed, it is plausible that studies with a relatively higher proportion of women may produce relatively higher effect sizes. Knowledge of such candidate cognitive mechanisms that may partly account for gender differences in depression, and explicit memory bias in particular could eventually help to ameliorate gender disparities in future research and treatment.

Age Group

Beck’s cognitive model proposes that dysfunctional schemas are formed during childhood and that these dysfunctional schemas are critically involved in the occurrence of depression during adulthood (A. T. Beck & Haigh, 2014; Clark et al., 1999). Schema-driven biases in information processing are proposed to represent a cognitive vulnerability marker that is relatively stable across the lifespan (Dozois & Dobson, 2001a; Goldstein et al., 2015). Consequently, mood-congruent biases should emerge regardless of age and occur in both minors (<18 years) and adults. However, compared to the literature on adults, a relatively small number of studies have investigated explicit LTM bias in samples of children and adolescents. Though several studies have observed a memory bias in children and adolescents with depression (Asl et al., 2015; Bishop et al., 2004; Connolly et al., 2016; Dujardin et al., 2014), other studies have failed to find differences between currently depressed or nondepressed minors (Alloy et al., 2012; Dalgleish et al., 2003; Hughes et al., 1990; Reid et al., 2006). Clarifying whether explicit LTM bias occurs in minors and adults could cast light on whether this cognitive factor represents a mechanism that operates across the lifespan.

Clinical Status of Depression

Cognitive theories of depression postulate that the magnitude of a memory bias differs depending on the clinical status of depression (Clark et al., 1999; Ingram, 1984). Researching differences between groups of individuals with depression, studies have recruited samples of patients with diagnosed major depression, undiagnosed individuals with self-reported elevated levels of depressive symptoms (often referred to as “subclinical depression”), and individuals remitted from major depression. In the face of more severe and impairing symptoms, theorists have argued that clinical depression is qualitatively different from subclinical symptom levels of depression (Ingram & Siegle, 2009). Hence, individuals with a diagnosed major depressive disorder are expected to display more severe explicit LTM bias than undiagnosed individuals with elevated depressive symptoms. Theoretical models also predict that LTM bias is not reversible with improvements in depressive symptoms but instead is a vulnerability factor that persists beyond the depressive episode resulting in a memory bias for even those remitted from depression (Clark et al., 1999; Ingram, 1984). This LTM bias in remitted depression may represent either a stable process that operates continuously or a latent process that is activated by stress or negative mood (Just et al., 2001), though it remains unclear whether there are differences in the magnitude of explicit LTM bias between groups with remitted, subclinical, and clinical forms of depression.

Though there has been little consideration of how comorbidity affects studies of explicit memory bias in depression, considering comorbid conditions is important due to the high co-occurrence between depression and other forms of psychopathology such as anxiety (Brown et al., 2001). Because other mental health conditions (e.g., social anxiety, panic disorder) may be characterized by a different pattern of cognitive biases (e.g., emotional biases at earlier stages of information processing; Armstrong & Olatunji, 2012; J. M. G. Williams et al., 1988, 1997), comorbidity may alter the pattern of biases that are expected to occur in depression. There is also
emerging evidence that negative memory bias operates across various mental disorders (Duyser et al., 2020; Herrera et al., 2017; Vrijsen et al., 2017). This evidence suggests explicit memory bias may play a mechanistic role in the development of co-occurrence between disorders and be unaffected by comorbidity between symptoms of different conditions.

**Status of Control Participants**

In studying explicit LTM bias, research has compared memory performance of participants with remitted, subclinical, and clinical depression with performance of various types of “control” participants. Studies have recruited individuals as “controls” with minimal self-reported levels of depressive symptoms using cutoff scores on established questionnaires, individuals who have never experienced a depressive episode or any mental disorder, and individuals who are currently not suffering from clinical depression. These types of “healthy control” participants may differ concerning their history of depression, current depressive symptom levels, or preferential processing of positive information (i.e., a positive explicit LTM bias). It is, therefore, possible that the nature of the control participants in samples modulates the strength of the overall effect size. Specifically, more extreme contrasts, such as comparing a depression group with a group of never-depressed or never-disordered individuals, may inflate the estimated population effect size (Fisher et al., 2020). However, research has yet to explore whether effect sizes of explicit LTM bias differ across samples including different types of control participants.

**What Are Boundary Conditions of Explicit Memory Bias in Depression?**

**Depth of Processing**

The depth of information processing has been theorized to modulate later memory performance (Craik, 2002; Craik & Lockhart, 1972). Conceptual processing (e.g., analysis of meaning, inference, and implication), as opposed to perceptual processing (e.g., analysis of stimulus features such as surface, form, color, brightness), is expected to result in higher levels of subsequent remembering. Because depression is marked by increased elaboration on negative material (Gotlib & Joormann, 2010; Ingram, 1984; J. M. G. Williams et al., 1988, 1997), a bias in explicit LTM is expected to occur following conceptual processing of emotional material but not following perceptual processing of presented stimuli.

A special case of conceptual processing is self-referential processing. Theoretical models hypothesize that a depression-linked bias in memory is pronounced for information that is relevant to a person’s negative schemas (Clark et al., 1999) or memory networks (Ingram, 1984). Investigating this hypothesis, studies have varied encoding task instructions requiring participants to process the presented stimuli in a self-referential or non-self-referential manner. During self-referential processing, participants process information in a manner that is relevant to their own character. For example, in the self-referential encoding task, participants make categorical decisions as to whether a presented adjective is self-descriptive or not (Connolly et al., 2016; Dainer-Best et al., 2017; Iacoviole et al., 2014). Prior research suggests that explicit memory bias mainly occurs following self-referential encoding (Wisco, 2009), but it is unclear whether explicit memory bias is elicited only following self-referential encoding. Because the depth of processing is a central feature in cognitive theories of depression, it is important to examine whether self-referential processing of stimuli enhances memory bias and if this bias can also occur when stimuli are processed in a perceptual or not self-referential conceptual manner.

**Intentionality of Encoding**

Studies have also varied the intention to learn during encoding to examine effects on mood-congruent memory bias in depression by employing incidental and intentional encoding tasks. In incidental encoding tasks, participants are not instructed to remember the stimuli presented during a cover encoding task such as rating the valence of the words. This cover encoding task is then followed by an unexpected memory test. In intentional encoding tasks, participants receive explicit instructions to memorize to-be-remembered information and are informed that there will be a memory test following encoding. Intentional encoding of emotional material typically enhances subsequent memory for encoded material (Ruiz-Caballero & González, 1994), but it is unclear whether incidental versus intentional encoding produces similar or different depression-linked bias in explicit LTM.

**Explicit Memory Task**

Four primary memory paradigms have been used to examine explicit LTM bias in depression, namely free-recall, cued-recall, forced-recall, and recognition tests. In the free-recall task, participants are asked to remember previously encoded items in any order (Reilly-Harrington et al., 1999; Romero et al., 2016). In a cued-recall task, participants are instructed to complete word stems with stimuli from the encoding task (Ilsley et al., 1995). A forced-recall task requires participants to remember a certain number of encoded stimuli and even guess stimuli when they could not remember any additional content (Murray et al., 1999). Finally, in recognition tasks, participants make old or new recognition decisions for individually presented items that are drawn from a set of learned stimuli intermixed with distractor stimuli (Ridout et al., 2003, 2009). During both recall and recognition tasks, participants may consciously retrieve contextual details that were associated with the item at the time of encoding. The difference between recall and recognition is that participants generate a response during recall tasks, whereas they select a response from given options during recognition tasks. Understanding differences between memory bias for recall and recognition tasks may help to characterize the nature of explicit LTM bias in depression.

**Type of Stimuli**

Experimental research on memory bias in depression has used a variety of stimulus materials during encoding tasks, including words (Dalglish et al., 2001; Holt et al., 2016), sentences (Everaert et al., 2013; Wenzlaff et al., 2002), stories (Bishop et al., 2004, p. 2004; Hughes et al., 1990), visual scenes (Harrington et al., 2018; Ramponi et al., 2010; Sears et al., 2011), and facial expressions (Gilboa-Schechtman et al., 2002; Wells & Beever, 2010). Most studies on depression have used verbal stimuli rather than more ecologically valid nonverbal stimuli such as visual scenes or pictures of facial expressions. It is currently unclear whether the strength of
explicit LTM bias differs for verbal or nonverbal stimuli. An examination of potential moderation effects would help determine whether the type of stimuli moderates the overall effect size to ascertain the generalizability of explicit LTM bias across various stimulus types.

Other Variables

The following procedural study features were examined in the current meta-analysis for exploratory purposes: stimulus presentation modality, number of stimuli presented during the encoding task, the number of stimulus repetitions during encoding, the presentation duration during encoding, the retention interval (i.e., the time between the encoding and retrieval phase), and the time allowed for recollection during the memory task.

The Present Review

This study aimed to provide a timely and comprehensive quantitative summary of almost 5 decades of research on emotional bias in explicit LTM in depression using multilevel meta-analytic techniques. The first objective was to assess the overall effect size of explicit LTM bias in depression. The second objective was to identify which sample characteristics, memory bias characteristics, and memory task features moderate the overall effect size. The moderator analyses aimed to address three core questions: (a) How does explicit memory bias occur in depression? (b) Who exhibits an explicit memory bias? And (c), what are boundary conditions of explicit memory bias in depression?

To achieve the study objectives, two separate meta-analyses on explicit LTM bias were conducted. The first meta-analysis examined naturalistic variations of explicit LTM bias in samples of individuals with remitted, subclinical, and clinical depression. The second meta-analysis investigated explicit LTM bias in remitted depression under naturalistic conditions and under conditions of induced negative mood or stress to determine whether explicit memory bias in remitted depression represents a vulnerability marker activated by stress or negative mood. An exhaustive examination of explicit LTM bias in various samples of depression is needed to draw empirically informed conclusions about the strength of the relation between depression and explicit memory bias, informing future theory, research, and treatments targeting explicit memory bias in depression.

Meta-Analysis I

Method

Identification of Studies

Studies were identified through complementary search strategies. First, studies were collected through comprehensive searches of electronic databases APA PsycINFO, Embase, ISI Web of Science (Science Citation Index Expanded, Social Science Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Conference Proceedings Citation Index—Science, Conference Proceedings Citation Index—Social Science & Humanities, Book Citation Index—Science, Book Citation Index—Social Science & Humanities), and PubMed through August 2021. To maximize coverage of the relevant studies, the following search string was entered into the databases: (depress* OR “mood disorder”* OR “affective disorder”*) AND (encod* OR memor* OR remember* OR mnemonic* OR recall OR recog* OR recollect* OR schema* OR retriev*) AND (self-referent* OR mood-congruen* OR bias* OR autobiograph*). Second, unpublished studies were searched in ProQuest Dissertations and Theses using the same search string as for published work. Finally, reference lists of review articles and meta-analyses were hand searched to ensure that as many relevant studies as possible were considered for inclusion.

Eligibility Criteria

Each article was assessed for relevance based on the following eligibility criteria used to guide the selection of studies for this meta-analysis.

1. The study concerned original empirical research with human participants reported in a full text and written in or translated to the English language.
2. The study investigated explicit LTM. Explicit LTM refers to long-term storage that represents knowledge (e.g., factual information, autobiographical events) in a consciously accessible manner that can be accessed via recall or recognition. Therefore, data from studies using free, cued, or forced-recall and recognition tasks were considered for inclusion. Studies employing paradigms to examine working memory (a temporary and limited-capacity store) or implicit LTM (nonconsciously accessible knowledge) were beyond the scope of this meta-analysis.
3. The study investigated LTM performance in terms of recall or recognition accuracy. Investigations of memory specificity, memory accessibility, memory intrusions, false memory, or memory intensity were not considered in this meta-analysis.
4. The study presented emotional (i.e., positive or negative) stimuli in the LTM task. Studies were allowed to present verbal (e.g., words) and/or nonverbal (e.g., images) emotional stimuli. Studies were excluded when they utilized ambiguous stimuli or neutral stimuli that acquired an emotional value during the study through experimental manipulation.
5. The study involved experimental procedures to control stimulus encoding. To minimize the possibility of encoding biases accounting for explicit LTM bias, studies had to involve an encoding phase when participants were presented with the to-be-remembered stimuli under controlled experimental conditions, followed by a retrieval phase probing memory for these stimuli. Studies were required to sufficiently describe the number of positive and/or negative stimuli that were presented as well as the encoding conditions (e.g., incidental vs. intentional) and instructions (e.g., conceptual vs. perceptual). Studies were excluded if they examined LTM for past life events for which the frequency of occurrence was unknown.
6. The study included at least one sample of participants with major depressive disorder or persistent depressive...
disorder, participants with a depressive disorder and comorbid anxiety disorder, participants remitted from a depressive disorder, or participants reporting elevated levels of depressive symptoms. The clinical status of major depressive disorder or persistent depressive disorder had to be assessed through clinical interview (e.g., Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; First et al., 2015)). The term clinical depression will be used to refer to participants currently meeting the diagnostic criteria for major depressive disorder or persistent depressive disorder according to the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013), Research Diagnostic Criteria (Kendler et al., 2010), or International Classification of Diseases (World Health Organization, 2018).

When clinical depression coexisted along with another disorder, the status was registered as “comorbid depression.” The status of remitted depression had to be determined through clinical interview or validated questionnaires (Zimmerman et al., 2004) to ensure that criteria for clinical depression were met in the past and not currently. The presence of elevated or “subclinical” levels of depressive symptoms in undiagnosed individuals had to be measured by validated questionnaires measuring depressive symptom severity levels (e.g., Beck Depression Inventory; A. T. Beck et al., 1996). Studies were excluded when the depressive disorder was not the principal diagnosis, participants suffered from mania or bipolar disorder, or when participant recruitment was guided by other variables than depression levels (e.g., addiction, anxiety, psychosis).

7. The study concerned natural variations in depression, depressive symptom levels, and emotional bias in explicit LTM. Studies employing procedures or interventions to alter participants’ mood state and/or memory bias (e.g., mood induction, cognitive training, pharmacological or psychological treatment) were excluded unless premanipulation data on explicit LTM bias and depression were reported.

8. The study sample consisted of children, adolescents, or adults. Sample recruitment was allowed from inpatient, outpatient, student/convenience, or community populations. Studies were excluded if the sample included individuals with known brain injury, neurological diseases, or persons with learning or developmental disorders.

9. The study utilized a categorical study design to compare explicit LTM bias in one or more depressed samples (e.g., elevated depressive symptoms, clinical, remitted depression) and a control group. In addition, studies were considered if they utilized a dimensional study design to examine explicit LTM bias along the continuum of depressive symptom severity.

10. The study reported sufficient statistics for the computation of effect sizes. If a study did not report sufficient information, the authors were contacted to provide the data required for inclusion. Studies were excluded if data necessary for inclusion were not retrieved.

**Selection of Studies**

Figure 1 outlines the literature search and winnowing process. A total of 12,273 records were identified through electronic database searches and 217 records were identified through other sources. After removal of duplicates, article titles and abstracts of 6,665 records were screened for relevance to this meta-analytic review. Record titles needed to refer to depression as well as memory. Abstracts needed to mention the use of one or more explicit LTM tasks with emotional stimulus materials. This reduced the number of relevant records to 518. The full text of the remaining 518 reports was read and assessed for eligibility according to the inclusion and exclusion criteria. This further reduced the number of relevant reports to 141. The reasons for the exclusion of 377 reports are detailed in Figure 1. A total of 141 reports described findings for 154 independent studies (N = 13,684) and 695 contrasts were included.1

The first and second authors conducted the selection process independently to ensure its reliability. Both raters judged the relevance of all records based on the outlined criteria. The interrater agreement was excellent (κ = .989). Cases of disagreement were solved through discussion until consensus was obtained.

**Data Coding Procedure**

A standardized coding system was applied to structure the coding process of each study. Coding was conducted by two independent raters. The agreement between raters was κ = .997. Disagreements were resolved by discussion until consensus was reached. Table 1 presents the characteristics of the studies included in this meta-analysis.

We coded the following study characteristics: author information, year of publication, publication status (published, unpublished), study design (categorical, dimensional), each article’s region of origin (country code, studies from the United States vs. outside the United States), and methodological quality. Regarding sample characteristics, we coded total sample/group size(s), the sample’s gender composition (proportion of female participants in the study), age group (minors, adults), clinical status (subclinical, clinical, remitted, comorbid),2 and control group (nondepressed, never-depressed, minimal depressive symptom levels). As for memory task and stimulus characteristics, we coded the stimulus type (verbal, nonverbal),3 stimulus modality (visual, auditory), encoding task (incidental, intentional), encoding depth (conceptual, perceptual), self-referent encoding (yes, no), number of to-be-encoded

---

1 This study considered 11 out of 14 studies that were included in a prior meta-analysis (Matt et al., 1992) and eight out of 11 studies included in a previous systematic review (Bogie et al., 2019). The other studies considered by these prior reviews did not offer the necessary statistics to calculate an effect size and were excluded.

2 Only eight studies recruited samples consisting of individuals with comorbid depression. Because this is below the recommended 10 studies for a characteristic (Harer et al., 2021; Higgins et al., 2022), it was not possible to consider comorbidity in the moderator analysis.

3 Studies have used a variety of verbal stimulus materials, including words, sentences, and scenarios. However, it was not possible to examine differences between the verbal stimulus types in the moderator analyses because of the small number of studies using scenario descriptions as stimuli in a memory task.
Methodological Quality

The methodological quality of included studies was assessed using a rating scale based on Downs and Black’s checklist for measuring quality (Downs & Black, 1998). This rating scale for nonrandomized study designs was recently adapted for use in meta-analytic research on cognitive processes in depression (Everaert et al., 2017; Gamble et al., 2019). The scale consisted of 21 items assessing the quality of several methodological components, namely reporting, external validity, bias in the measurement (internal validity), confounding in the selection of study participants (internal validity), and power of the study. Criteria were rated on a 2-point scale (1 = yes, 0 = no, 0 = unable to determine), except for one item in the reporting subscale which was scored on a 3-point scale (2 = yes, 1 = partially, 0 = no). An average score was computed for each methodological component (except for power, which consisted of one item). Higher scores indicate better quality. The rating scale used in this study is provided in Supplemental Material 1.5

Meta-Analytic Procedure

The Hedges’ g statistic was used as the measure of effect size. Hedges’ g is commonly used to represent the standardized difference between means and corrects for bias in small sample sizes (Hedges & Olkin, 1985), yielding an unbiased estimate of the

Figure 1
Flow of the Study Search and Winnowing Process for Meta-Analysis 1

Note. See the online article for the color version of this figure.

4 Only one study used a forced-recall task and only six studies used a cued-recall task. Because this is below the recommended 10 studies for a characteristic (Higgins et al., 2022), it was not possible to consider forced- and cued-recall tasks in the moderator analysis.

5 In using the version of the Downs and Black’s checklist for cognitive processes in depression, we deviated from the methodological quality measure described in the preregistration protocol (see Supplemental Material 2). This decision was informed by the better psychometric properties of the Downs and Black’s checklist (Downs & Black, 1998) and less ambiguous items of the rating scale (yes vs. no/unable to determine) as compared to the other method (that rates items on a scale with good/fair/poor as anchors). This decision was made prior to rating the methodological quality of individual studies.

stimuli, number of stimulus repetitions during encoding, stimulus presentation duration during encoding (in seconds), duration of the retention interval (in minutes), type of memory task (free-recall, forced-recall, cued-recall, recognition),4 time allowed for memory recollection (in minutes), and the number of memory test cycles. Finally, regarding the memory bias index characteristics, we coded the valence (positive, negative, threat) and operational definition (between-subject, within-subject).
population standardized mean difference (Cumming, 2012). For studies with a categorical design, Hedges’ $g$ was computed using means, standard deviations, and group sizes. Between-group effect sizes were calculated by dividing the mean difference between the depressed and the control group by the pooled standard deviation (Borenstein et al., 2009). Within-group effect sizes were computed by dividing the mean difference between two memory measurements within the depressed group (negative vs. neutral, positive vs. neutral, negative vs. positive) by the average standard deviation of both repeated measurements (Lakens, 2013). For dimensional study designs, we calculated Hedges’ $g$ from Fisher’s $r$-to-$z$ transformed correlation coefficients representing the association between depressive symptom levels and explicit memory bias to empirically aggregate and synthesize results from different study designs. This conversion assumes that continuous data used to compute the correlation coefficients have a bivariate normal distribution and that the two groups are created by dichotomizing one of the two variables (i.e., the variable representing individual differences in depressive symptom levels; Borenstein et al., 2009, p. 48). When studies reported the number of study participants with low or minimal versus elevated depressive symptom levels, a formula for equally sized or unequally sized groups was used (Harrer et al., 2021). When means and standard deviations or correlation coefficients were not reported, Hedges’ $g$ was computed based on other statistics reported in the article: Cohen’s $d$, one-way $F$ statistics, means and standard errors, independent $t$ tests, and exact $p$ values. Hedges’ $g$ values were computed such that a positive effect size indicated greater mood congruence (i.e., more negative or less positive) in explicit LTM bias. Hedges’ $g$ values between 0.2–0.5, 0.5–0.8, and ≥0.8 refer to small, medium, and large effect sizes, respectively (Cohen, 1988). In addition to the summary estimate and 95% confidence intervals (CIs), 95% prediction intervals (PIs) are reported. The 95% PI estimates locate ranges within which true effects are to be expected for 95% of similar studies that might be conducted in the future (IntHout et al., 2016).

Three-level random-effects models with restricted maximum-likelihood estimation were fitted in R Version 4.1.2 (R Core Team, 2021) using the rma.mv function from the metafor package (Viechtbauer, 2010). The three-level meta-analytic model is a strong method to account for dependencies among effect sizes (e.g., multiple effect sizes from one sample) during meta-analytic pooling (Cheung, 2014; Van den Noortgate et al., 2015). Most studies reported multiple memory bias contrasts and were included in the present study (e.g., within- and between-subject comparisons, recall of positive and negative memories, memory bias scores obtained from free-recall and recognition tasks). The coef_test() function from the clubSandwich package (Pustejovsky, 2022) was used to apply cluster-robust standard errors to the fitted models and handle dependent sampling errors of the effect size estimates within studies. The three-level meta-analytic model allows the estimation of three sources of variance: sampling variance of the extracted effect sizes (Level 1), variance between effect sizes extracted from the same study (Level 2), and variance between effect sizes extracted

### Table 1

**Overall Characteristics of Studies in Meta-Analyses I and II**

<table>
<thead>
<tr>
<th>Study variable</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of publication</td>
<td>$M = 2004$, $SD = 11$</td>
<td>$M = 2010$, $SD = 8$</td>
</tr>
<tr>
<td>Publication status (Publ/NPubl)</td>
<td>$k = 137/k = 17$</td>
<td>$k = 20/k = 1$</td>
</tr>
<tr>
<td>Study design (Cat/Dim/Sing)</td>
<td>$k = 126/k = 26/k = 2$</td>
<td>$k = 21/k = 0/k = 0$</td>
</tr>
<tr>
<td><strong>Sample characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sample/group size(s)</td>
<td>$M = 89$, $SD = 125$</td>
<td>$M = 74$, $SD = 69$</td>
</tr>
<tr>
<td>Proportion of female participants</td>
<td>$M = 65$, $SD = 18$</td>
<td>$M = 72$, $SD = 17$</td>
</tr>
<tr>
<td>Age group (minors/adults)</td>
<td>$k = 25/k = 129$</td>
<td>$k = 2/k = 19$</td>
</tr>
<tr>
<td>Clinical status (CD/SD/RD/D+)</td>
<td>$k = 77/k = 66/k = 15/k = 8$</td>
<td>$k = 77/k = 66/k = 15/k = 8$</td>
</tr>
<tr>
<td>Control group (MS/NeD/NoD)</td>
<td>$k = 83/k = 49/k = 16$</td>
<td>$k = 2/k = 16/k = 2$</td>
</tr>
<tr>
<td><strong>Memory task and stimulus characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mood induction (Y/N)</td>
<td>$k = 129/k = 23/k = 3$</td>
<td>$k = 7/k = 14$</td>
</tr>
<tr>
<td>Stimulus type (verbal/nonverbal/mix)</td>
<td>$k = 126/k = 18/k = 10$</td>
<td>$k = 19/k = 2/k = 1$</td>
</tr>
<tr>
<td>Stimulus modality (visual/auditory/mix)</td>
<td>$k = 108/k = 45/k = 1$</td>
<td>$k = 14/k = 7/k = 0$</td>
</tr>
<tr>
<td>Encoding task (incidental/intentional/mix)</td>
<td>$k = 119/k = 29$</td>
<td>$k = 16/k = 6$</td>
</tr>
<tr>
<td>Encoding depth (conceptual/perceptual)</td>
<td>$k = 67/k = 104$</td>
<td>$k = 8/k = 13$</td>
</tr>
<tr>
<td>Self-referent encoding (Y/N)</td>
<td>$M = 43$, $SD = 28$</td>
<td>$M = 64$, $SD = 30$</td>
</tr>
<tr>
<td>Number of to-be-encoded stimuli</td>
<td>$k = 1$, $SD = 1$</td>
<td>$M = 1$, $SD = 0$</td>
</tr>
<tr>
<td>Number of stimulus repetitions</td>
<td>$k = 9$, $SD = 27$</td>
<td>$M = 3$, $SD = 3$</td>
</tr>
<tr>
<td>Stimulus presentation duration(s)</td>
<td>$M = 511$, $SD = 2.447$</td>
<td>$M = 1.685$, $SD = 3.863$</td>
</tr>
<tr>
<td>Duration of retention interval (min)</td>
<td>$k = 113/k = 60/k = 3/k = 42$</td>
<td>$k = 11/k = 0/k = 3/k = 8$</td>
</tr>
<tr>
<td>Type of memory task (CRec/FRec/FoRec/FoRec)</td>
<td>$M = 4$, $SD = 3$</td>
<td>$M = 3$, $SD = 1$</td>
</tr>
<tr>
<td>Time allowed for recollection (min)</td>
<td>$M = 2$, $SD = 3$</td>
<td>$M = 2$, $SD = 4$</td>
</tr>
<tr>
<td>Memory bias index characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence (Neg/Pos/Thr)</td>
<td>$k = 149/k = 131/k = 26$</td>
<td>$k = 19/k = 20/k = 2$</td>
</tr>
<tr>
<td>Operational definition (between subjects/within subjects)</td>
<td>$k = 147/k = 113$</td>
<td>$k = 21/k = 16$</td>
</tr>
</tbody>
</table>

**Note.** $k =$ number of studies; Publ = published; NPubl = not published; Cat = categorical; Dim = dimensional; Sing = single group; CD = clinical depression; SD = subclinical depression; RD = remitted depression; D+ = depression with comorbid disorder; MS = minimal symptoms; NeD = never depressed; NoD = nondepressed; Y = yes; N = no; FRec = free recall; CRec = cued recall; FoRec = forced recall; Recog = recognition; Neg = negative; Pos = positive; Thr = threat.
from different studies (Level 3; Cheung, 2014; Hox et al., 2017). To determine whether the within-study variance (Level 2) and between-study variance (Level 3) was significant, two separate log-likelihood-ratio tests compared the deviance scores of the full model and models excluding the Level 2 or Level 3 variance parameter.

Sources of within-study and between-study variation were examined through planned moderator analyses. Separate three-level random-effects models were fitted with a restricted maximum-likelihood model and the Knapp–Hartung method for each candidate moderator variable. Moderator analyses were conducted if each category contained a minimum of five studies given that parameter estimates are poor when the number of studies is very small.

The risk of bias was assessed in five ways. First, components of methodological quality (i.e., reporting, external validity, bias, confounding, and power) in the meta-analysis were examined as moderators of the overall effect size. Second, the magnitude of effect sizes in published and unpublished studies was compared through moderator analysis. Third, regions of article origin (studies from the United States vs. outside the United States) were examined in moderator analyses to explore whether the magnitude of effect sizes is similar in diverse contexts. Fourth, funnel plots were examined to determine whether small studies systematically generated larger point estimates than larger studies (Jin et al., 2015). A funnel plot depicts the effect size of each study on the x-axis and their standard error on the y-axis. In the absence of a bias, the plot has a funnel-like shape with studies dispersed and distributed symmetrically around the pooled effect size (Borman & Grigg, 2009). Funnel plot asymmetry was further evaluated through Egger’s regression test (Egger et al., 1997). To this end, a meta-regression model (using cluster-robust variance estimation) was tested that included the inverse standard errors of the effect size estimates as a moderator and the observed effect sizes divided by their standard error as an outcome. Finally, sensitivity analyses as proposed by Mathur and VanderWeele (2020) were conducted to examine bias due to selective publication and reporting. These sensitivity analyses quantify the amount of publication bias that would be required to attenuate the observed point estimate to null. The analyses for the robust clustered specification assumed that publication bias would favor studies showing a memory bias in depression. The significance funnel plot was examined to detect the extent to which point estimates of the nonaffirmative studies (i.e., studies with a nonsignificant or negative estimate) are systematically smaller than the entire set of point estimates. The amount of publication bias required to attenuate g to the null was estimated. Relatively large ratios indicate that a meta-analysis is relatively robust to publication bias (Mathur & VanderWeele, 2020).

**Results**

**Characteristics of the Studies**

Table 1 summarizes the characteristics of the included studies. A list of the included studies with their characteristics is provided in Supplemental Material 5. The studies originated from various regions in the world including Australia, Belgium, Canada, China, Denmark, France, Germany, Israel, Iran, Ireland, New Zealand, Norway, Poland, Serbia, South Korea, Spain, The Netherlands, Switzerland, the United Kingdom, and the United States. However, most studies were conducted in Canada (11%), the United Kingdom (23%), and the United States (39%). Almost all studies (96%) were conducted in high-income countries as listed by the World Bank. On average, four memory bias contrasts (SD = 3, range: 1–16) were extracted from the same participants in the independent samples or groups in a study. Several studies provided 10 or more effect size estimates (Bradley & Mathews, 1988; Hammen et al., 1986; Pincus et al., 1995; Roat, 2006; Sears et al., 2011; Sellstrom, 1989; Shestyk, 2006; Shestyk & Deldin, 2010; Tarsia et al., 2003; Timbremont & Braet, 2004; van der Laat, 2004). Most studies were published records, employed a categorical study design, recruited adult samples, and studied individuals with current (clinical or subclinical) depression. Sample sizes ranged from 10 to 1,015 totaling 13,684 participants. The proportion of female participants ranged from 0% to 100% with an average of 65%. With regard to memory task properties, most studies used an incidental encoding task and required participants to process the information at a conceptual level. The number of studies using a self-referential encoding task was lower than the number of studies using a conceptual encoding task that was not self-referential. Most studies displayed stimuli visually and presented verbal stimulus materials. Memory was most frequently tested using a free-recall task or recognition task. Finally, most studies examined memory bias for positive or negative emotional content and studied memory bias for threatening content less frequently.

**Overall Effect Size and Heterogeneity Among Effect Sizes**

The overall effect size was estimated by fitting an intercept-only three-level random-effects model using data of 695 contrasts from 154 independent samples. Results showed a pooled effect size estimate of $g = 0.28$, 95% CI [0.20, 0.37], 95% PI [-1.21, 1.78], $p < .001$. Statistical outlier analysis identified nine contrasts from four independent samples (Callahan et al., 2017; Dozios & Dobson, 2001b; Moulds et al., 2007; Sloan et al., 2001) with standardized residual values exceeding 3 (Aguinis et al., 2013; Viechtbauer & Cheung, 2010). These outlying effect size estimates may impact the precision and interpretation of the overall effect size and were excluded from subsequent analyses.

Refitting the intercept-only model (686 contrasts, 153 studies) yielded a small overall effect size of $g = 0.24$, 95% CI [0.18, 0.30], 95% PI [-1.05, 1.54], $p < .001$. There was significant variation between effect sizes within studies, $\sigma^2 = 0.40$, $\chi^2(1) = 1708.89$, $p < .001$, as well as between studies, $\sigma^2 = 0.03$, $\chi^2(1) = 7.62$, $p = .006$. The estimated sampling error variance was 0.05. Of the total variance in effect sizes, 7% was accounted for by variance between studies, 82% by variance between effect sizes of the same study, and 11% by random sampling variance. The caterpillar plot in Figure 2a depicts the magnitude of effect sizes included in the meta-analysis with corresponding 95% CIs.
**Moderator Analyses**

Characteristics of the sample, memory bias, and task features were examined as potential sources of the variation between effect sizes. Table 2 presents the results of the moderator analyses.

**How Does Explicit Memory Bias Occur in Depression?** The effect of emotional valence was significant, $F(2, 571) = 5.91, p = .003$. The effect sizes of memory bias in depression were positive and significant for negative and positive material but nonsignificant for threat-related material. Of note, the effect size for positive material was larger than the effect size for negative material, estimate difference: $g = 0.17, 95\% \text{ CI } [0.019, 0.328], p = .028$. The effect of operational definition was also significant, $F(1, 684) = 9.73, p = .002$. Effect sizes for both within-subject and between-subject comparisons were positive and significant, though between-subject comparisons yielded larger effects than within-subject comparisons, estimated difference: $g = 0.17, 95\% \text{ CI } [0.07, 0.27], p < .001$.

**Who Exhibits an Explicit Memory Bias?** The effect of gender composition was not significant, $F(1, 631) = 0.18, p = .670$. The effect of age group was not significant, $F(1, 684) = 2.34, p = .126$. Effect sizes were positive and significant in samples of minors and adults. Furthermore, the clinical status of depression showed no significant effect, $F(2, 650) = 1.97, p = .141$. Effect sizes for samples including individuals with clinical and subclinical depression were positive and statistically significant. Of note, the effect size for samples including individuals with remitted depression was not reliably different from zero. Finally, there was a nonsignificant effect for control group, $F(2, 672) = 0.20, p = .815$. Effect sizes were positive and significant for samples using participants with minimal symptom levels, nondepressed status, or never-depressed status as a comparison group.

**What Are Boundary Conditions of Explicit Memory Bias in Depression?** The effect of intentionality of encoding was not significant, $F(1, 681) = 2.80, p = .095$. Effect sizes for studies using incidental and intentional encoding tasks were positive and significant. The effect of depth of processing was significant, $F(2, 671) = 5.55, p = .004$. Positive and significant effect sizes were found for all processing conditions. The effect size for self-referential conceptual encoding was larger than the effect size for not self-referential conceptual encoding, estimated difference: $g = 0.20, 95\% \text{ CI } [0.056, 0.335], p = .007$, and perceptual encoding, estimated difference: $g = 0.22, 95\% \text{ CI } [0.037, 0.397], p = .019$. There was no significant difference between effect sizes from not self-referential conceptual and perceptual encoding, estimated difference: $g = 0.02, 95\% \text{ CI } [-0.124, 0.167], p = .167$. Moreover, the effect of explicit memory task was significant, $F(1, 651) = 6.90, p = .009$. Effect sizes were significantly larger when emotional memory was examined using a free-recall task compared to a recognition task, estimated difference: $g = 0.19, 95\% \text{ CI } [0.059, 0.311], p = .005$. The effect sizes for free-recall and recognition tasks were significantly different from zero. Finally, the effect of type of stimuli was significant, $F(1, 672) = 4.56, p = .033$. The effect size for memory bias in depression was stronger in conditions with verbal stimuli than in conditions with nonverbal stimuli, estimated difference: $g = 0.18, 95\% \text{ CI } [0.01, 0.35], p = .036$. The effect size for conditions presenting nonverbal stimuli was not reliably different from zero.

The exploratory analyses revealed that the effect of stimulus modality did not reach the .05 threshold of statistical significance, $F(1, 635) = 3.27, p = .071$. A significant effect size was found for conditions presenting stimuli auditorily, estimated difference: $g = 0.17, 95\% \text{ CI } [0.01, 0.32]$.
Table 2
The Effect of Moderators on the Relation Between Explicit LTM Bias and Depression for the Main and Follow-Up Meta-Analysis

<table>
<thead>
<tr>
<th>Meta-analysis I</th>
<th>Moderator</th>
<th>Level</th>
<th>n</th>
<th>k</th>
<th>g (SE)</th>
<th>95% PI</th>
<th>t(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who exhibits an explicit memory bias?</td>
<td>Gender composition</td>
<td>Adults</td>
<td>633</td>
<td>143</td>
<td>0.001 (0.002)</td>
<td>[−1.11, 1.57]</td>
<td>0.430 (28.30)</td>
<td>.670</td>
</tr>
<tr>
<td></td>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minors</td>
<td></td>
<td>577</td>
<td>127</td>
<td>0.219 (0.032)</td>
<td>[−1.07, 1.51]</td>
<td>6.890 (103.00)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Clinical status</td>
<td>Clinical</td>
<td></td>
<td></td>
<td>322</td>
<td>76</td>
<td>0.233 (0.039)</td>
<td>[−1.07, 1.54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subclinical</td>
<td></td>
<td></td>
<td>275</td>
<td>66</td>
<td>0.278 (0.053)</td>
<td>[−1.03, 1.58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Remitted</td>
<td></td>
<td></td>
<td>56</td>
<td>15</td>
<td>0.050 (0.099)</td>
<td>[−1.27, 1.37]</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>Minimal</td>
<td></td>
<td></td>
<td>344</td>
<td>83</td>
<td>0.263 (0.048)</td>
<td>[−1.04, 1.57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Never depressed</td>
<td></td>
<td></td>
<td>247</td>
<td>49</td>
<td>0.226 (0.053)</td>
<td>[−1.08, 1.53]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-depressed</td>
<td></td>
<td></td>
<td>84</td>
<td>15</td>
<td>0.211 (0.056)</td>
<td>[−1.10, 1.53]</td>
</tr>
<tr>
<td>What are boundary conditions of explicit memory bias in depression?</td>
<td>Intentionality of encoding</td>
<td>Incidental</td>
<td></td>
<td></td>
<td>451</td>
<td>107</td>
<td>0.272 (0.040)</td>
<td>[−1.02, 1.56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intentional</td>
<td></td>
<td></td>
<td>232</td>
<td>45</td>
<td>0.161 (0.047)</td>
<td>[−1.13, 1.45]</td>
</tr>
<tr>
<td></td>
<td>Depth of processinga</td>
<td>Perceptual</td>
<td></td>
<td></td>
<td>148</td>
<td>33</td>
<td>0.150 (0.061)</td>
<td>[−1.14, 1.44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conceptual-aRef</td>
<td></td>
<td></td>
<td>286</td>
<td>68</td>
<td>0.172 (0.040)</td>
<td>[−1.12, 1.46]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conceptual-Sref</td>
<td></td>
<td></td>
<td>240</td>
<td>67</td>
<td>0.367 (0.061)</td>
<td>[−0.92, 1.66]</td>
</tr>
<tr>
<td></td>
<td>Explicit memory taska</td>
<td>Free recall</td>
<td></td>
<td></td>
<td>483</td>
<td>112</td>
<td>0.289 (0.040)</td>
<td>[−1.02, 1.59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Recognition</td>
<td></td>
<td></td>
<td>170</td>
<td>42</td>
<td>0.104 (0.049)</td>
<td>[−1.20, 1.41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonverbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Verbal</td>
<td></td>
<td></td>
<td>562</td>
<td>128</td>
<td>0.272 (0.036)</td>
<td>[−1.03, 1.57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visual</td>
<td></td>
<td></td>
<td>291</td>
<td>125</td>
<td>0.273 (0.038)</td>
<td>[−1.05, 1.60]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Auditory</td>
<td></td>
<td></td>
<td>108</td>
<td>18</td>
<td>0.105 (0.062)</td>
<td>[−1.23, 1.44]</td>
</tr>
<tr>
<td></td>
<td>Number of stimuli</td>
<td></td>
<td></td>
<td></td>
<td>661</td>
<td>145</td>
<td>−0.002 (0.001)</td>
<td>[−1.06, 1.55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stimulus repetitions</td>
<td></td>
<td></td>
<td>679</td>
<td>151</td>
<td>−0.050 (0.028)</td>
<td>[−1.06, 1.54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presentation duration</td>
<td></td>
<td></td>
<td>396</td>
<td>87</td>
<td>−0.001 (0.001)</td>
<td>[−0.88, 1.35]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Retention interval</td>
<td></td>
<td></td>
<td>630</td>
<td>142</td>
<td>−0.000 (0.000)</td>
<td>[−1.09, 1.52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Recollection time</td>
<td></td>
<td></td>
<td>313</td>
<td>74</td>
<td>0.000 (0.008)</td>
<td>[0.87, 1.46]</td>
</tr>
<tr>
<td></td>
<td>Memory test cyclesa</td>
<td></td>
<td></td>
<td></td>
<td>685</td>
<td>152</td>
<td>−0.034 (0.010)</td>
<td>[−1.05, 1.52]</td>
</tr>
<tr>
<td></td>
<td>Emotional valencea</td>
<td>Negative</td>
<td></td>
<td></td>
<td>278</td>
<td>141</td>
<td>0.201 (0.053)</td>
<td>[−1.11, 1.51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positive</td>
<td></td>
<td></td>
<td>254</td>
<td>124</td>
<td>0.375 (0.051)</td>
<td>[−0.93, 1.68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Threat</td>
<td></td>
<td></td>
<td>42</td>
<td>27</td>
<td>0.050 (0.167)</td>
<td>[−1.28, 1.38]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within subject</td>
<td></td>
<td></td>
<td>239</td>
<td>110</td>
<td>0.128 (0.039)</td>
<td>[−1.56, 1.41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Between subject</td>
<td></td>
<td></td>
<td>447</td>
<td>148</td>
<td>0.301 (0.039)</td>
<td>[−0.98, 1.58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unpublished</td>
<td></td>
<td></td>
<td>107</td>
<td>17</td>
<td>0.074 (0.054)</td>
<td>[−1.22, 1.57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Published</td>
<td></td>
<td></td>
<td>579</td>
<td>136</td>
<td>0.266 (0.035)</td>
<td>[−1.02, 1.56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reporting</td>
<td></td>
<td></td>
<td>686</td>
<td>153</td>
<td>−0.308 (0.248)</td>
<td>[−1.05, 1.54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>External validity</td>
<td></td>
<td></td>
<td>686</td>
<td>153</td>
<td>−0.011 (0.080)</td>
<td>[−1.05, 1.54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td></td>
<td></td>
<td>686</td>
<td>153</td>
<td>0.291 (0.284)</td>
<td>[−1.05, 1.54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confounding</td>
<td></td>
<td></td>
<td>686</td>
<td>153</td>
<td>−0.241 (0.133)</td>
<td>[−1.05, 1.55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Power</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No/unclear</td>
<td></td>
<td></td>
<td>628</td>
<td>139</td>
<td>0.224 (0.030)</td>
<td>[−1.07, 1.57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td>58</td>
<td>14</td>
<td>0.389 (0.168)</td>
<td>[−0.92, 1.69]</td>
</tr>
<tr>
<td></td>
<td>Article’s region of origin</td>
<td>Studies from the United States</td>
<td></td>
<td></td>
<td>273</td>
<td>55</td>
<td>0.185 (0.051)</td>
<td>[−1.11, 1.48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Studies outside the United States</td>
<td></td>
<td></td>
<td>413</td>
<td>98</td>
<td>0.274 (0.043)</td>
<td>[−1.02, 1.57]</td>
</tr>
</tbody>
</table>

Note.  LTM = long-term memory; n = number of contrasts; k = number of studies; g = mean effect size (Hedges’ g); SE = standard error; 95% PI = 95% prediction interval.

A significant moderator at the .05 threshold of statistical significance.
p = .034. The effect size for conditions with auditory stimulus presentation was not reliably different from zero. Effects of stimulus repetitions, $F(1, 677) = 1.28, p = .258$, presentation duration, $F(1, 394) = 0.33, p = .568$, length of the retention interval, $F(1, 628) = 1.46, p = .228$, recollection time, $F(1, 311) = 0.000, p = .993$, and the number of stimuli, $F(1, 659) = 1.96, p = .162$, did not moderate the overall effect size. The effect of memory test cycles was significant, $F(1, 683) = 9.400, p = .002$. Smaller effect sizes for explicit memory bias in depression occurred when studies utilized a greater number of test cycles during the memory task.

Follow-up analyses crossed levels of the significant moderators to identify conditions for which larger effect sizes occurred. Results from these exploratory analyses are provided in Supplemental Material 6. Examining depth of encoding—explicit memory task combinations revealed that self-referential encoding followed by free recall produced larger effect sizes than non-self-referential encoding followed by free recall, estimated difference: $g = 0.19, 95\% CI [0.04, 0.35], p = .013$, non-self-referential encoding followed by recognition, estimated difference: $g = 0.27, 95\% CI [0.10, 0.43], p = .002$, perceptual encoding followed by free recall, estimated difference: $g = 0.22, 95\% CI [0.03, 0.41], p = .026$, and perceptual encoding followed by recognition, estimated difference: $g = 0.32, 95\% CI [0.10, 0.55], p = .006$. All conditions yielded effect sizes that were significantly larger than zero, except for perceptual encoding followed by a recognition task.

Examining depth of encoding by type of stimulus combinations, results showed that verbal self-referential processing yielded larger effect sizes than verbal non-self-referential processing, estimated difference: $g = 0.19, 95\% CI [0.05, 0.34], p = .010$, nonverbal non-self-referential processing, estimated difference: $g = 0.25, 95\% CI [0.04, 0.45], p = .019$, and verbal perceptual processing, estimated difference: $g = 0.19, 95\% CI [0.01, 0.37], p = .044$. All conditions but the nonverbal non-self-referential encoding conditions produced effect sizes that were significantly larger than zero.

Finally, explicit memory task by type of stimulus combinations was examined. Results indicated that verbal free-recall procedures produced a larger effect size than verbal recognition, estimated difference: $g = 0.22, 95\% CI [0.07, 0.36], p = .044$. All effect sizes were significantly larger than zero, except for recognition tasks produced using verbal stimuli.

**Explained Heterogeneity in Effect Sizes.** A three-level model including all significant moderator effects was fitted to examine variance explained by the moderators. This model considered only effect sizes that have data for all moderators (537 contrasts, 135 studies). As in prior work (Houben et al., 2015), the variance components of the model with all moderators were compared to the intercept-only model that included the same effect sizes. Results showed that the within-study variance was reduced by 3.03\%, $\chi^2(1) = 1215.14, p < .001$, but significant variation remained between effect sizes within studies, $\sigma^2 = 0.43, \chi^2(1) = 1215.14, p < .001$. The between-study variance was reduced by 100\%, $\chi^2(1) = 0.00, p = 1.000$, explaining all variation between effect sizes between studies.

**Assessment of Bias**

Several approaches were used to determine the robustness of the estimated effects in this meta-analysis. First, several aspects of the methodological quality of studies were examined as a moderator. The methodological components of reporting, $F(1, 684) = 1.40, p = .237$, external validity, $F(1, 684) = 0.02, p = .897$, bias in measurement, $F(1, 684) = 0.59, p = .445$, confounding, $F(1, 684) = 2.78, p = .096$, and power, $F(1, 684) = 2.44, p = .119$, did not moderate the strength of memory bias in depression. Exploratory follow-up analyses were conducted for the subset of studies with a total quality score higher than the mean. The intercept-only model (290 contrasts, 61 studies) yielded a small overall effect size of $g = 0.20, 95\% CI [0.10, 0.30], p < .001, 95\% PI [−0.97, 1.37] that was not substantially different from the original pooled effect size estimate of $g = 0.24$. There was significant variation between effect sizes within studies, $\sigma^2 = 0.28, \chi^2(1) = 791.98, p < .001$, and between studies, $\sigma^2 = 0.07, \chi^2(1) = 18.36, p < .001$. These observations suggest that the included effect sizes and their heterogeneity were not biased by different components of methodological quality. Quality ratings for each criterion and methodological component are presented in Table 3. Some marked weaknesses emerged. Few studies reported a power analysis to justify the sample size, actual probability values for all statistical tests, or withdrawals and dropouts. In addition, studies did not consistently document whether invited and recruited research participants were representative of the population and if groups differed concerning principal confounders.

Second, the magnitude of effect sizes in published versus unpublished studies was compared through moderator analysis. Results showed that the effect of publication status was significant, $F(1, 684) = 4.57, p = .033$. Stronger effect sizes were reported in published studies compared to unpublished studies, estimated difference: $g = 0.19, 95\% CI [0.06, 0.33], p = .009$. This suggests that studies with null or negative findings were less likely to be published.

Third, moderator analysis compared the magnitude of effect sizes from studies that were conducted inside the United States versus outside the United States. Results showed that the effect of article origin was not significant, $F(1, 684) = 1.91, p = .168$. Effect sizes were positive and significant for both studies from the United States and studies conducted outside the United States.

Fourth, the funnel plot (see Figure 3) was inspected to investigate the relationship between the effect sizes and standard errors of the included studies. The plot suggests a funnel-like shape. Testing funnel plot asymmetry, the metaregression model did not produce a significant intercept, $b_0 = 0.82, (4.50) = 1.48, p = .146$. This suggests no small-study effects (Harrer et al., 2021).

Finally, the significance funnel plot (Figure 4) suggests that the worst-case estimate of only nonaffirmative studies, $g = −0.09, 95\% CI [−0.16, −0.03], p = .003$, was close to null and in the opposite direction of the original pooled estimate of all studies. Numerical sensitivity analyses under robust clustered specification indicated that affirmative results would need to be at least 3.87 times more likely to be published than nonaffirmative results to attenuate the pooled point estimate to the null. This indicates that a substantial publication bias would be required to explain away the meta-analytic pooled estimate (Mathur & VanderWeele, 2020).

**Summary**

This meta-analysis estimated the overall effect size of naturalistic variations in explicit memory bias in depression and tested moderators that could explain variability in research findings. Supporting theoretical predictions (A. T. Beck & Haigh, 2014; Bower, 1981; Hertel, 2004a; Ingram, 1984; J. M. G. Williams et al., 1997),
results revealed a small overall effect size indicating that depression is marked by an explicit memory bias. Moderator analysis revealed that this bias is expressed as a within-subject and between-subject difference, such that individuals with depression recall relatively more negative or less positive than neutral material, and that this cognitive profile differs from emotional memory retrieval in individuals without depression. Effect sizes were significantly larger for the lack of a positive memory bias in depression than for a negative memory bias. The effect of threat-related bias was not significant.

Moderator analyses further showed that the depth but not the intentionality of encoding moderated the overall effect size. Effect sizes were larger for tasks using self-referential encoding than for tasks requiring participants to engage in conceptual, non-self-referential encoding, or perceptual encoding. Results showed that effect sizes were significantly larger for free-recall tasks than for recognition tasks. Regarding the type of stimuli, experimental conditions using verbal stimuli produced larger effect sizes than experimental conditions using nonverbal stimuli.

Exploratory analyses revealed that stimulus modality, stimulus repetitions, presentation duration, length of the retention interval, and recollection time did not moderate the overall effect size. There was a marginal effect for the number of stimuli and a significant effect for memory test cycles. This suggests that smaller effect sizes for memory bias may emerge when experimental task conditions utilized a greater number of stimuli and test cycles during the memory task.

Significant effect sizes were found for studies in minors and adults, suggesting that explicit memory bias occurs across different age groups. The status of the control group did not moderate the overall effect size. While statistically significant effect sizes were found for studies of participants with clinical depression and studies of participants with self-reported elevated (subclinical) depressive symptoms, a nonsignificant effect size emerged for studies of participants who remitted from depression. The latter finding that there is no consistent support for an explicit memory bias in remitted depression challenges the hypothesis that explicit memory bias persists beyond a depressive episode and represents a cognitive vulnerability factor of depression. However, it should be noted that this meta-analysis excluded studies using mood or stress manipulation to activate a mood-congruent bias of explicit LTM, thereby only including studies examining naturalistic variations in memory bias. It is plausible that explicit memory bias in remitted depression represents a vulnerability marker that is activated by stress or negative mood, a hypothesis that aligns with most cognitive models (A. T. Beck & Haigh, 2014; Bower, 1981; Ingram, 1984). To examine this possibility, a follow-up meta-analysis was conducted focusing on explicit memory bias in remitted depression under both naturalistic and induced mood conditions or stressors.

### Table 3
**Ratings for the Adapted “Checklist for Measuring Quality”**

<table>
<thead>
<tr>
<th>Item</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Hypotheses, aims, objectives clearly identified?</td>
<td>0.99</td>
<td>0.08</td>
</tr>
<tr>
<td>2. Primary outcomes clearly described?</td>
<td>0.99</td>
<td>0.08</td>
</tr>
<tr>
<td>3. Participant characteristics clearly described?</td>
<td>0.83</td>
<td>0.38</td>
</tr>
<tr>
<td>4. Study procedure clearly described?</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5. Consideration of principal confounders?</td>
<td>0.62</td>
<td>0.49</td>
</tr>
<tr>
<td>6. Main findings clearly described?</td>
<td>0.97</td>
<td>0.18</td>
</tr>
<tr>
<td>7. Estimates for random variability of main outcomes reported?</td>
<td>0.88</td>
<td>0.32</td>
</tr>
<tr>
<td>8. Withdrawals and dropouts reported?</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>9. Actual probability values reported?</td>
<td>0.28</td>
<td>0.45</td>
</tr>
</tbody>
</table>

**Internal validity—bias**

<table>
<thead>
<tr>
<th>Item</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Invited subjects representative?</td>
<td>0.41</td>
<td>0.49</td>
</tr>
<tr>
<td>11. Participating subjects representative?</td>
<td>0.19</td>
<td>0.39</td>
</tr>
</tbody>
</table>

**Internal validity—confounding**

<table>
<thead>
<tr>
<th>Item</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Data dredging made clear?</td>
<td>0.90</td>
<td>0.30</td>
</tr>
<tr>
<td>13. Appropriate use of statistical tests?</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>14. Methods used to assess depression valid and reliable?</td>
<td>0.99</td>
<td>0.08</td>
</tr>
<tr>
<td>15. Main memory outcome measures unbiased and correct?</td>
<td>0.98</td>
<td>0.14</td>
</tr>
</tbody>
</table>

**External validity**

<table>
<thead>
<tr>
<th>Item</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. Participants recruited from same population?</td>
<td>0.56</td>
<td>0.50</td>
</tr>
<tr>
<td>17. Participants recruited within the same time window?</td>
<td>0.85</td>
<td>0.36</td>
</tr>
<tr>
<td>18. Differences between groups on principal confounders? Were depression scores related to any of these demographic variables?</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>19. Adequate adjustment for confounding?</td>
<td>0.79</td>
<td>0.41</td>
</tr>
<tr>
<td>20. Were missing data handled appropriately?</td>
<td>0.92</td>
<td>0.27</td>
</tr>
</tbody>
</table>

**Power**

<table>
<thead>
<tr>
<th>Item</th>
<th>Meta-analysis I</th>
<th>Meta-analysis II</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. Power analysis reported or adequate sample size</td>
<td>0.10</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Note.** All items have a maximum score of 1.00.
Eligibility Criteria

Eligibility criteria were identical to criteria applied in the main meta-analysis, except for two modifications:

1. The study included at least one sample of patients who remitted or recovered from a depressive disorder. The status of remitted depression had to be determined through clinical interview and/or validated questionnaires to ensure that criteria for clinical depression were met in the past and not currently.

2. The study concerned natural variations in emotional bias of explicit LTM and/or manipulated emotional LTM bias using a mood or stress induction. Studies employing cognitive training, pharmacological, or psychological treatment to alter memory bias were excluded unless premanipulation data on explicit LTM bias and depression were reported.

Selection of Studies

Figure 5 details the literature search and winnowing process. A total of 387 records were identified. After removal of duplicates, article titles and abstracts of 272 records were screened. This reduced the number of relevant records to 47. The full-text reports were read and assessed for eligibility according to the inclusion and exclusion criteria. This further reduced the number of relevant
reports to 21. The reasons for the exclusion of 26 reports are detailed in Figure 5. A total of 21 reports describing findings for 21 independent studies (N = 1,561) and 83 contrasts were included. Two independent raters judged the relevance of all records based on the outlined criteria. The interrater agreement was excellent (κ = .91). Cases of disagreement were solved through discussion until full agreement was obtained.

Data Coding Procedure

A standardized coding system was applied to every study. In addition to the characteristics coded for the main meta-analysis, we coded whether the study used a mood induction procedure (yes, no). Coding was conducted by two independent raters. The agreement between raters was κ = .913. Disagreements were resolved by
discussion until consensus was reached. Supplemental Material 7 presents the characteristics of the studies included in this follow-up meta-analysis.

**Methodological Quality Assessment**

The methodological quality of included studies was again assessed using a rating scale based on Downs and Black’s checklist for measuring quality (Downs & Black, 1998) as in the first meta-analysis.

**Meta-Analytic Procedure**

The same procedure for effect size calculation, data synthesis, and assessment of bias was followed as in Meta-analysis I. This meta-analysis examined mood induction (yes, no) as a moderator in addition to moderating variables related to the nature of the memory bias (emotional valence, operational definition). Publication bias was determined through funnel plot inspection, analysis of methodological quality components as moderators, and sensitivity analysis for publication bias.

**Transparency and Openness**

PRISMA 2020 reporting guidelines were followed for the final report (Page et al., 2021). The PRISMA 2020 checklist is provided in Supplemental Material 4. The data, R code, and output of the analyses are available on the Open Science Framework (https://osf.io/wseb4; Everaert et al., 2022).

**Results**

**Characteristics of the Studies**

Table 1 summarizes the characteristics of the included studies. A list of the included studies with their characteristics is provided in Supplemental Material 7. The studies originated from various regions in the world including Belgium, Canada, Germany, Spain, The Netherlands, the United Kingdom, and the United States. Most studies were conducted in Canada (14%) and the United States (51%). All studies were conducted in high-income countries as listed by the World Bank. On average, four relevant memory bias contrasts ($SD = 2$, range: 1–8) were obtained from the same participants from the independent samples or groups in a study. Two studies provided eight effect size estimates (Ramel et al., 2007;
Overall Effect Size and Heterogeneity Among Effect Sizes

Fitting an intercept-only three-level random-effects model using data of 83 contrasts from 21 independent samples yielded an overall effect size of $g = 0.09$, $95\%$ CI $[-0.10, 0.28]$, $95\%$ PI $[-1.53, 1.78]$, $p = .338$. Three contrasts from one study (Sears et al., 2011) had standardized residual values exceeding 3. These outlying effect size estimates may impact the precision and interpretation of the overall effect size and were excluded in subsequent analyses (Aguinis et al., 2013; Viechtbauer & Cheung, 2010). Refitting the intercept-only model (80 contrasts, 21 studies) yielded a pooled effect size that was not significantly different from zero, $g = 0.13$, $95\%$ CI $[-0.05, 0.31]$, $95\%$ PI $[-1.06, 1.32]$, $p = .136$. The caterpillar plot in Figure 2b depicts the effect sizes with corresponding $95\%$ Cs. There was significant variation between effect sizes within studies, $\sigma^2 = 0.32$, $\hat{\chi}^2(1) = 107.51$, $p < .001$, but not between studies, $\sigma^2 = 0.03$, $\hat{\chi}^2(1) = 0.69$, $p = .407$. The estimated sampling error variance was 0.06. Of the total variance in effect sizes, 7% was accounted for by variance between studies, 78% by variance between effect sizes of the same study, and 15% by random sampling variance. Therefore, moderator analysis examined potential sources of the variation between effect sizes. Table 2 presents the results of the moderator analyses.

Moderator Analyses

The effect of the moderator mood or stress induction was not statistically significant, $F(1, 78) = 1.45$, $p = .232$. Inspecting effect sizes for two levels of the moderator, results showed a statistically significant pooled effect size for conditions with mood or stress induction but not for conditions without a mood or stress induction procedure. The difference between both effect sizes was not significant, $g = 0.21$, $95\%$ CI $[-0.082, 0.501]$, $p = .139$. However, caution is necessary in interpreting this finding because only seven studies implemented a mood induction procedure. This is below the recommended 10 studies for a level of a moderator (Harrer et al., 2021; Higgins et al., 2022).

Moreover, the effect of emotional valence was not significant, $F(1, 59) = 0.07$, $p = .790$. Effect sizes for both positive and negative memory bias were not statistically different from zero. However, the effect of operational definition approached the .05 threshold for statistical significance, $F(1, 78) = 3.53$, $p = .064$. A significant effect size occurred for between-subject but not within-subject comparisons, estimated difference: $g = -0.28$, $95\%$ CI $[-0.57, 0.02]$, $p = .064$. Together, these moderators accounted for some variability between effect sizes within studies. The within-study variance was reduced by 2.75%, $\hat{\chi}^2(1) = 55.40$, $p < .001$, and between-study variance by 1.55%, $\hat{\chi}^2(1) = 2.95$, $p = .086$.

Note that all additional moderators tested in Meta-analysis I were also examined post hoc in Meta-analysis II when there was a sufficient number of studies at each level of the moderator. Except for the number of test cycles, no other moderators were statistically significant. The results of the post hoc moderator analysis are provided in Supplemental Material 8.

Assessment of Bias

Moderator analysis showed that the methodological components of reporting, $F(1, 78) = 0.15$, $p = .702$, external validity, $F(1, 78) = 0.10$, $p = .751$, bias, $F(1, 78) = 0.08$, $p = .774$, and confounding, $F(1, 78) = 0.33$, $p = .565$, did not modulate the strength of memory bias in remitted depression. The moderating role of statistical power could not be evaluated because only five contrasts from one sufficiently powered study were observed. Exploratory follow-up analyses were conducted for the subset of studies with a total quality score higher than the mean. The intercept-only model (38 contrasts, 10 studies) yielded an overall effect size of $g = 0.17$, $95\%$ CI $[-0.16, 0.50]$, $p = .122$, $95\%$ PI $[-1.75, 2.09]$. This pooled effect size was similar to the original pooled effect size estimate of $g = 0.13$. There was significant variation between effect sizes within studies, $\sigma^2 = 0.87$, $\hat{\chi}^2(1) = 129.71$, $p < .001$, but not between studies, $\sigma^2 = 0.00$, $\hat{\chi}^2(1) = 0.00$, $p = 1.000$. These findings suggest that the included effect sizes and their heterogeneity were not biased by differences in characteristics of methodological quality. The average quality ratings for studies in Meta-analysis II are presented in Table 3.

Second, moderator analysis compared the magnitude of effect sizes from studies that were conducted inside versus outside the United States. Results showed that the effect of the article’s origin was not significant, $F(1, 78) = 3.18$, $p = .080$. Effect sizes were positive and significant for non-U.S. studies but nonsignificant for studies conducted in the United States. However, caution is necessary in interpreting the latter finding because of the small number of studies (<10 studies) included.

Third, the funnel plot (see Figure 3b) was inspected to investigate the relationship between the effect sizes and standard errors of the included studies. The plot showed that studies were dispersed and distributed symmetrically around the pooled effect size, forming a funnel-like shape. The metaregression model did not show a significant intercept, $b_0 = 0.30$, $t(5.42) = 0.55$, $p = .606$. The nonsignificant intercept suggests that the meta-analysis did not miss out on small-study effects (Harrer et al., 2021).

Finally, the significance funnel plot (Figure 4) suggests that the worst-case estimate of only nonaffirmative studies, $g = -0.16$, $95\%$ CI $[-0.30, -0.02]$, $p = .025$, was in the opposite direction of the original pooled estimate of all studies. Numerical sensitivity analyses under robust clustered specification indicated that affirmative results would need to be 1.97 times more likely to be published than nonaffirmative results for publication bias to explain away the meta-analytic pooled point estimate completely. A moderate publication bias would be required to shift the point estimate of 0.13 to the null (Mathur & VanderWeele, 2020).
Summary

This meta-analysis examined the presence of explicit memory bias in remitted depression under naturalistic conditions as well as mood or stress induction. As in the first meta-analysis, results showed a small effect size that was not significantly different from zero. However, there was considerable heterogeneity among effect sizes. Examining sources of this heterogeneity, results showed that the use of a mood induction did not significantly moderate the overall effect size. While the effect sizes for samples of participants with remitted depression were not reliably different from zero under naturalistic conditions, the pooled effect size for conditions with mood or stress induction was statistically significant. Valence was not a significant moderator. Effect sizes for both positive and negative memory bias were not statistically different from zero. However, the operational definition approached statistical significance. Effect sizes for between-subject comparisons but not within-subject comparisons were significant.

Discussion

Consistent with predictions by major cognitive models of depression (A. T. Beck & Haigh, 2014; Bower, 1981; Hertel, 2004a; Ingram, 1984; J. M. G. Williams et al., 1997), literature reviews (Bogie et al., 2019; LeMoult & Gotlib, 2019; Wisco, 2009), and a previous meta-analysis conducted 30 years ago (Matt et al., 1992), Meta-analysis I revealed a small but significant overall effect size supporting a naturalistic explicit memory bias in depression. Meta-analysis II provided some support for an explicit memory bias under mood induction in remitted depression. In both meta-analyses, there was significant variation between effect sizes within or between studies. A priori identified moderators were tested to determine the sources of this heterogeneity and answer three questions about explicit memory bias in depression (see Table 2).

How Does Explicit Memory Bias Occur in Depression?

Meta-analysis I demonstrated that the strength of the between-subject bias was significantly larger than the within-subject bias, but both were statistically significant. This observation helps to characterize naturalistic explicit memory bias in depression and integrates previous findings focusing on either a within-subject bias (Matt et al., 1992) or a between-subject bias (Bogie et al., 2019). Instead of singularly operationalizing explicit memory bias, future studies should adopt a multilevel perspective to investigate within-versus-between-subject causes and effects of this bias in depression.

The findings from Meta-analysis I also support that depression is marked by both a negative memory bias and the lack of a positive memory bias (A. T. Beck & Haigh, 2014; Bower, 1981; Hertel, 2004a; Ingram, 1984; J. M. G. Williams et al., 1988, 1997). That the small-to-moderate effect size for a reduced positive memory bias was larger than the small effect for a negative bias is notable in light of the bulk of research focusing on how negative information is processed (Everaert et al., 2012; Gotlib & Joormann, 2010; LeMoult & Gotlib, 2019; Mathews & MacLeod, 2005). This difference in magnitude underscores the importance of considering positive emotion deficits in cognition–emotion interactions in depression. A reduced positive memory bias may help to account for decreased positive affect and anhedonia as hallmark symptoms of depression (American Psychiatric Association, 2013) and emotion regulation difficulties in response to positive emotions (e.g., dampening) that maintain low positive affect (Vanderlind et al., 2020, 2021). In addition, this finding has implications for emerging cognitive training interventions targeting memory bias to treat symptoms of depression (Becker et al., 2015; Visser et al., 2020). That is, memory interventions may benefit from techniques facilitating the encoding and retrieval of positive events rather than a focus on negative experiences.

Who Exhibits an Explicit Memory Bias?

The finding that studies including a higher proportion of women did not produce larger effect sizes is remarkable in light of recent work suggesting a stronger association between depression symptoms and memory bias among adolescent females (Bone et al., 2021). Some have argued that this stronger association may occur only during developmental periods characterized by elevated risk for psychopathology such as adolescence (Kessler et al., 2005). Further research is needed to explore potential gender differences in explicit memory bias across the lifespan.

Examining potential age-related differences, Meta-analysis I indicated that minors and adults with depression display a bias that is similar in magnitude. While consistent with conceptualizations of schema-driven memory bias as a cognitive marker formed during childhood (A. T. Beck & Haigh, 2014; Clark et al., 1999; Dozois & Dobson, 2001a; Goldstein et al., 2013), this evidence provides only indirect support for the notion of memory bias as a stable cognitive factor. Longitudinal research is needed to track individuals during transitions across the lifespan to uncover when explicit memory bias in depression emerges and stabilizes.

The finding that individuals with clinical depression may not exhibit a more severe explicit memory bias than undiagnosed individuals with elevated symptoms suggests that clinical depression may not be qualitatively different from its nonclinical forms concerning explicit memory bias. Dimensional approaches examining explicit memory bias, as opposed to clinical cutoffs, may thus be more appropriate to study vulnerability to depression (Gibb et al., 2004; Sanislow, 2020). Unfortunately, because of the limited number of studies, this meta-analysis could not test whether the presence of other mental health conditions alters the strength of the effect size of explicit memory bias in depression. Examining the role of comorbidity provides an interesting direction for future meta-analytic reviews.

Perhaps most remarkably, Meta-analysis I did not provide support for an explicit memory bias in remitted depression under naturalistic conditions. Follow-up Meta-analysis II also considered studies using mood induction and its results suggested that a memory bias in depression may only occur following a mood or stress induction. This observation confirms predictions by cognitive models that explicit memory bias persists beyond the depressive episode (Clark et al., 1999; Ingram, 1984) as a latent process activated by stress or negative mood (Just et al., 2001). This finding is also consistent with a prior meta-analysis showing a lingering deficit in LTM capacity following recovery (Semkovska et al., 2019). Importantly, this observation should be interpreted cautiously because only seven studies in remitted samples used mood induction. This number of studies is below the recommended 10 studies for a characteristic to be included in moderator analyses.
(Harrer et al., 2021; Higgins et al., 2022) and precludes strong conclusions about whether memory biases can be induced through mood or stress induction in remitted depression.

Meta-analysis II also showed that remitted depression may feature a between-subject bias in the absence of a within-subject bias. This finding suggests that remitted individuals do not have an explicit memory bias but individuals with no symptoms or history of depression remember emotional material in a biased manner. Because of the relatively low number of included studies, Meta-analysis II lacked power to further explore conditions for which effect sizes are larger (e.g., mood induction following self-referential processing for verbal stimulus materials). Future research could manipulate experimental conditions to study whether configurations of the moderators examined here reveal a robust explicit memory bias in remitted depression.

What Are Boundary Conditions of Explicit Memory Bias in Depression?

Findings within this domain of inquiry may be especially relevant toward informing the design of future experiments, namely consideration of the level of processing, the use of free-recall or recognition encoding tasks, and types of stimuli. The finding that self-referential encoding is linked to a stronger explicit memory bias in depression is consistent with models postulating that a depression-linked bias in explicit memory is triggered by information that is relevant to a person’s negative schemas (Clark et al., 1999) or memory networks (Ingram, 1984). This observation also adds to the growing body of research indicating that self-referential processing is a critical cognitive marker of depression (Alloy et al., 1997; Connolly et al., 2016; Dainer-Best et al., 2017). However, studies using a conceptual encoding that was not self-referential did not yield larger effect sizes than studies using perceptual encoding conditions, which is inconsistent with other depth of processing accounts assuming that conceptual processing results in higher levels of subsequent remembering (Craik, 2002; Craik & Lockhart, 1972). The absence of differences is also at odds with enhanced elaboration models that assume a stronger explicit memory bias would occur following conceptual processing (Ingram, 1984; J. M. G. Williams et al., 1988, 1997). Taken together, this pattern of findings suggests that depth of processing during encoding of emotional material does not determine whether an explicit memory bias occurs, it only contributes to the magnitude of the resulting bias.

Meta-analysis I observed significant effects for incidental and intentional encoding tasks, but the difference between them was not statistically significant. Though intentional encoding of emotional material typically enhances subsequent memory for encoded material in previous work (Ruiz-Caballero & González, 1994), these results suggest that explicit memory bias in depression is relatively robust against variations in intentionality during encoding.

Factors operating during the retrieval phase may also impact explicit memory bias in depression. The fact that free recall yielded larger effects than recognition tasks suggests that they may have mechanistic differences even though both tasks are commonly regarded as explicit memory tasks. In recognition tasks, two retrieval processes may contribute to emotional memory task performance, namely recollection (i.e., retrieval of specific details about the prior occurrence of an item) and familiarity (i.e., a sense of having encountered an item without retrieving specific details; Everaert & Koster, 2015; Jermann et al., 2009). Unfortunately, most studies using recognition tasks did not distinguish between these two cognitive operations, thus small effects of implicit memory bias may have obfuscated explicit memory effects in recognition tasks (Gaddy & Ingram, 2014). In contrast, free-recall tasks seem to rely predominantly on controlled cognitive operations, requiring participants to engage in a self-initiated memory search process driven by internal representations in the absence of external retrieval cues (Polyn et al., 2009). This aspect of self-regulatory control during memory retrieval may be particularly problematic in depression (Strauman, 2017) and may explain stronger explicit memory bias on free-recall tasks.

The effect size for memory bias in depression was stronger for verbal stimuli than for nonverbal stimuli such as scenes and facial expressions, which are intrinsically more vivid, arousing, and emotion provoking than words (Bartoszek & Cervone, 2017; Kensinger & Schacter, 2006). This finding suggests that individuals with depression retrieve emotional memories from explicit memory in a verbal manner, but it is also possible that nonverbal stimuli differ in their quality to reflect concerns relevant to depression, subsequently diminishing effects. Additionally, studies have typically used nonverbal stimuli within memory recognition paradigms that also produce smaller effects than free-recall paradigms. Thus, future research using visual stimuli could use them in free-recall tasks as well as assess stimuli quality in tapping depression-related concerns. Among several other moderators tested, only the number of memory test cycles yielded a significant effect such that smaller effect sizes occur when studies utilized a greater number of stimuli and test cycles. These observations indicate that explicit memory bias is relatively invariant to various memory task features.

In exploratory analyses crossing levels of these moderators, uncovered specific conditions for which larger effect sizes occurred (see also Supplemental Material 6). In particular, study conditions using self-referential encoding followed by free recall, verbal self-referential processing, or verbal free recall produced relatively larger effect sizes. Study conditions using perceptual encoding followed by a recognition task, nonverbal non-self-referential encoding, and verbal recognition tasks produced effect sizes that were not significantly different from zero. Future hypothesis-driven research should confirm whether these conditions produce a stronger explicit memory bias related to depression.

Assessment of Bias

The overall quality of studies was high and did not modulate the magnitude of the overall effects, yet some weaknesses provide directions for future work. First, studies should include a power analysis to justify the sample size, report precise probability values for all statistical tests, and transparently describe withdrawals and dropouts. Second, studies should elaborate on whether the invited and recruited research participants are representative of the targeted population. A considerable portion of research in psychological science is conducted in predominantly WEIRD samples (western, educated, industrialized, rich, democratic; Henrich et al., 2010) that typically lack demographic diversity. This limitation also applies to the current meta-analysis. Though the article’s region of origin did not moderate the overall effect size, most studies in Meta-analysis I and all studies in Meta-analysis II were conducted in high-income
countries. Because many groups of people seem underrepresented in research on memory bias in depression, it is unclear whether the current findings and implications generalize to more diverse populations. Finally, as in other areas of research on cognitive biases in depression (O’Connor et al., 2021), many studies on explicit memory biases in depression failed to report psychometric properties such as reliability of the different experimental paradigms used in the study (for notable exceptions, see Hjordt et al., 2020; Jensen et al., 2016). These three issues represent a threat to the transparency, replicability, and societal impact of research findings in this area.

Though funnel plots and funnel plot asymmetry tests from Meta-analyses I and II suggested that these did not miss out on small-study effects (see Figures 3 and 4), results revealed that publication status modulated the magnitude of the overall effect size in Meta-analysis I. A small but significant effect size was observed for published studies, whereas the effect size for unpublished studies did not significantly differ from zero. This suggests that research in this area may be plagued by a publication bias. Note that Meta-analysis II included only one unpublished study and publication status could not be tested as a moderator. Further examining bias due to selective publication and reporting in both meta-analyses, sensitivity analyses indicated that moderate (for Meta-analysis II) and substantial (for Meta-analysis I) publication bias would be required to explain away the meta-analytic pooled point estimate completely. Overall, this pattern suggests that the findings from both meta-analyses are relatively robust to publication bias.

Limitations

Several limitations should be acknowledged. First, it cannot be ruled out that the sample included here is not representative of all studies that have been conducted to investigate the relationship between depression and explicit memory bias. Though the risk of missing relevant studies was minimized, it was not possible to obtain necessary information for various studies published between 1970 and 1980. In addition, this review was based on English-language reports obtained from search strategies in databases that are predominantly filled with English-language reports. This may have introduced bias to the sample of studies. As a result, early research on explicit memory bias in depression and research published in local journals may be less well represented in this review. Future research should consider non-English sources and conduct multilingual searches to determine the robustness of explicit memory bias in depression across nations, cultures, and language groups.

Second, there was considerable heterogeneity of the pooled effect size that a priori identified moderators were not able to fully explain. The fact that only a very small portion of the heterogeneity could be explained is remarkable because key moderators derived from theory and methodology were considered by this meta-analytic study. This limitation highlights the need to identify further study-specific and individual-differences factors (e.g., the age of first onset of depression, number of past depressive episodes, and receipt of treatment) that moderate explicit memory bias in depression. In light of the high level of unexplained heterogeneity, the results of the present meta-analysis should be interpreted cautiously. As indicated by the PIs (see Table 2), the range into which effects of future studies may fall is broad.

Third, this study was limited in its ability to examine most moderators conjointly. A systematic approach of combining the 17 moderators in Meta-analysis I would result in 137 additional analyses. With such a high number of exploratory analyses, there is a considerable risk of obtaining spurious findings. Additionally, this study did not combine more than two moderator variables in any analyses because this may introduce imprecision of estimates by considering only a subset of studies (and effect size estimates) when crossing levels of moderators. Therefore, exploratory analyses were limited to crossing levels of pairs of two significant moderators to identify boundary conditions for which larger effect sizes occurred.

Finally, this meta-analysis mostly included data from studies examining explicit memory bias in depression at a single time point in highly controlled experimental conditions. These designs have limited the ecological validity in addition to obscuring the direction of the relationship between explicit memory bias and depression. The evidence for explicit memory bias in current depression does speak to its potential role as a vulnerability or contributing factor, but to test this would require multiwave longitudinal or experimental research manipulating memory bias to assess its influence on depressive symptoms. Some recent studies have used ecological momentary assessment (EMA) to study explicit memory bias in real-life contexts, reporting that an EMA-based explicit memory bias metric was significantly correlated with both depressive symptoms and rumination, as well as the free-recall index derived from the self-referential encoding task administered under experimental conditions (Vrijsen et al., 2021). Future research could embed EMA approaches within longitudinal or behavioral high-risk study designs to help elucidate the temporal relationship between explicit memory bias and the development of depressive symptoms, as well as circumvent limitations associated with remitted depressed designs.

Conclusion

This meta-analytic review provides a comprehensive examination of several decades of research on explicit memory bias in depression. The findings provide an updated benchmark of empirical research for the field and evidence for the presence of explicit memory bias in samples of minors and adults currently suffering from depressive symptoms. The findings from the presented meta-analyses support its theorized role as a vulnerability marker of depression but also highlight limitations to current knowledge, including the limited heterogeneity in effect sizes that was accounted by key theoretical and methodological moderators. Several directions for future research are outlined for next steps in this exciting research area.

References

References marked with an asterisk were included in the meta-analysis.


Explain the concept of explicit memory bias in depression and provide examples of research findings.
EXPLICIT MEMORY BIAS IN DEPRESSION 461


Received June 25, 2021
Revision received July 6, 2022
Accepted July 11, 2022

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!