UNIVERSITY OF CALGARY

The Stroop Task in Depression: A Meta-Analysis

by

Amanda M. Epp

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

PROGRAM IN CLINICAL PSYCHOLOGY

CALGARY, ALBERTA

SEPTEMBER, 2008

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "The Stroop Task in Depression: A Meta-Analysis" submitted by Amanda M. Epp in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, Dr. Keith Dobson, Department of Psychology

Dr. Candace Konnert, Department of Psychology

Dr. Penny Pexman, Department of Psychology

Dr. John Mueller, Department of Applied Psychology

September 8, 2008

Date

Abstract

Despite the extensive use of the Emotional Stroop task in depression, only qualitative reviews have been produced to date, and these reviews conclude that Stroop performance in depression is equivocal. The present meta-analysis addressed the need to quantitatively summarize the data. A thorough search of the literature was conducted and 38 published studies and unpublished doctoral dissertations were included in the analyses. The meta-analysis revealed large and robust depression-related Stroop effects. Although the effects did not reflect an emotion-congruent bias, they did distinguish among levels of depressive experience, where greater severity was associated with larger effect sizes. Moreover, the effects did not require priming procedures, longer stimulus exposure, or the presentation of self-relevant or disorder-congruent stimuli, to be obtained. Given these findings, further research on the Stroop task in depression is not necessary. Research using a more direct measure of depression-related attentional bias is recommended.

Acknowledgements

I would like to thank my supervisor Dr. Keith Dobson for all of his support, encouragement, and guidance throughout the completion of this project. I would like to extend appreciation to my thesis committee for their willingness to assist me with this process, and for their valuable feedback in the formative stages of this project. I would also like to thank Darcy and my parents, Sue and Helmut, who have always provided unending assistance, support, encouragement, patience, and faith in my abilities. My fellow graduate students and the research assistants in my lab have all been a great source of assistance and support over the past two years as well, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Running Head: THE STROOP TASK IN DEPRESSION

The Stroop Task in Depression: A Meta-Analysis

Amanda M. Epp

Department of Psychology

Program in Clinical Psychology

University of Calgary

The Stroop task in depression: A meta-analysis

The Stroop task (Stroop, 1935) has received considerable attention over the course of the last seventy years. MacLeod (1991) estimated that there were "more than 700 Stroop-related articles in the literature" (p. 163), and the numbers have increased since that time. Continued interest in the Stroop may be attributable to the widening breadth of applications for which the Stroop has been adapted, such as research on cognitive models of psychopathology. Meta-analytic reviews of Stroop performance can now be found for such diverse groups as eating disordered populations (Dobson & Dozois, 2004), addicted populations (Cox, Fadardi, & Pothos, 2006), anxious populations (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007), attention-deficit hyperactivity disordered populations (Lansbergen Kenemans, & van Engeland, 2007), and populations with schizophrenia (Szöke et al., 2008). Despite the extensive use of the Stroop task in depression, only qualitative reviews have been produced to date (Gotlib, Roberts, & Gilboa, 1996; Mogg & Bradley, 2005; Williams, Mathews, & MacLeod, 1996). These reviews conclude that Stroop performance in depression is equivocal. Thus, it is timely to conduct a meta-analytic review to ascertain the state of the science.

The Stroop Task

The original Stroop task (variably referred to in the literature as the Classic Stroop, or the Stroop Colour-Word Task) was developed in 1935 by J. Ridley Stroop. The task involves the presentation of colour words, in incongruously coloured ink (e.g., the word 'red' printed in blue ink). Participants are asked to name, as quickly as possible, the ink colour of each stimulus word, while attempting to ignore the meaning of the word

(Stroop, 1935). This attempt to suppress word meaning in order to name ink colour has reliably been shown to result in longer response latencies than those that result from colour naming congruent stimuli (e.g., the word 'red' printed in red ink; MacLeod, 1991). Response latency scores are the dependent variable in some Stroop studies, and are typically measured in milliseconds as the time between the onset of the stimulus and the participant's response. Interference scores are the dependent variable in other Stroop studies, where interference is measured as the difference in response latencies between incongruent and congruent stimuli, or between incongruent and coloured non-lexical stimuli (e.g., colour patches). The fact that incongruent stimuli produce longer response latencies than congruent or coloured non-lexical stimuli is referred to as the 'Stroop effect'.

The purpose of Stroop studies varies from an exploration of the theoretical account of the Stroop effect, to an examination of treatment outcomes, or individual differences that affect Stroop performance (MacLeod, 1991). The Stroop is viewed as a useful tool in psychopathology research, to examine the cognitive processes of the disorder under investigation (Williams et al., 1996). Specifically, it is presumed that disordered individuals are sensitive to and preoccupied by stimuli related to their concern (Williams et al., 1996), and the Stroop task is used to determine the existence and nature of this attentional bias. The original Stroop task has been modified into an "Emotional" or "Modified Stroop" for this research, by changing the content of the word stimuli from colours to affectively-laden themes related to the disorder under investigation (Stroop, 1935, MacLeod, 1991). The relative response latency associated with the delayed naming of disorder-related words, compared with neutral words, is assumed to reflect an

attentional bias for the disorder-related stimuli (Gotlib et al., 1996). Similarly, poorer Stroop performance among disordered participants compared with controls is assumed to reflect a disorder-specific attentional bias.

The Stroop Task in Depression

The importance of examining the nature of an attentional bias in depression is evident, given that a mood-congruent attentional bias has been postulated to play a role in the etiology and maintenance of the disorder (Dalgleish & Watts, 1990; Gotlib et al., 1996; Mogg & Bradley, 2005). Stroop studies assess the disruptive impact of depression on Stroop performance, as an index of the disruptive impact of attention allocation to negative thought patterns, while performing task-related activities (Segal, 1996). Such studies typically involve a comparison between a depressed and a control sample on depression-specific, negative, positive, and neutral stimuli. The results of these studies are often interpreted in light of Beck's cognitive theory of depression (Beck, Rush, Shaw, & Emery, 1979), and Bower's (1981) network theory of emotion.

Beck's model proposes that early life experiences form the basis for interconnected negative schemas about the self, the world, and the future which bias information processing (e.g., attention, encoding, retrieval; Dozois & Dobson, 2001). Schemas are the relatively stable and enduring internal templates that store, organize, integrate, and direct the processing of information (Beck et al., 1979; Segal & Swallow, 1994). Individuals who suffer from depression tend to interpret situations in a distorted manner, in line with their underlying negative schemas (Beck et al., 1979). Bower's (1981) work adds to Beck's conceptualization of depression by demonstrating the powerful association between emotion and cognition. His associative network theory

states that current emotional state influences associative processes, the interpretation of ambiguous situations, and the salience of congruent emotional material. The theory further predicts that emotion enhances the salience of mood-congruent material for selective attention and learning. Reciprocal associations between mood and thoughts occur due to activation of one or the other, through associative linkages. Thus, individuals are predicted to actively attend to material consistent with their feelings, and mood-congruent stimuli should be more salient than content that is not mood-congruent (see also Clark, Beck, & Alford, 1999). Longer reaction times to negative stimuli are thought to reflect the greater effort required to suppress the meaning of those highly accessible schema-congruent stimuli (Segal & Swallow, 1994).

Cognitive Mechanisms Underlying Depression-Related Bias on the Stroop Task

Although Beck's and Bower's models provide a framework for how the Emotional Stroop effect should manifest with depressed populations, they do not explain the underlying cognitive mechanisms that produce the Stroop effect. Several information processing theories have been proposed to explain these mechanisms. The differences between the theories relate primarily to whether interference occurs at an early encoding stage or at a later response stage of information processing. While debate about the most appropriate model of the Stroop effect continues, the most accepted model to date is the Parallel Distributed Process framework (PDP; Cohen, Dunbar, & McClelland, 1990). The PDP offers an intermediary explanation of the Stroop effect that encompasses both early and late phases of processing.

The PDP framework posits that there is a cognitive pathway for colour naming and another for word reading (Cohen et al., 1990). The two pathways each consist of

input units (representing colour or words), intermediate units, and output units (representing response to name the colour or read the word). The pathways are interconnected by these units at multiple levels, and thus the action of one may be disrupted or facilitated by the action of the other at any point, after sensory perception. According to Cohen et al. (1990), attention modulates the system, as it alters the responsiveness of the units according to whether the task demands that the colour be named or the word read. Speed and accuracy of a response depend on the strength of processing of each pathway, influenced by the adjacent pathway, and modulated by task demand. Williams et al. (1996) reviewed the evidence for this model in the anxiety and depression literature, and concluded that this model captures the data reasonably well.

Evidence for a Depression-Related Stroop Effect

Beck's and Bower's models predict that depression should be associated with an attentional bias for mood-congruent stimuli. With the Stroop task, this prediction suggests mood-congruent interference effects for depressed populations, compared with controls. On the other hand, several researchers have argued that depression is associated with biases in controlled or effortful processing, such as that involved in interpretation and memory, and not with early or relatively automatic processes such as attention (Hartlage, Alloy, Vázquez, & Dykman, 1993; Williams, Watts, MacLeod, & Mathews, 1988; Mathews & MacLeod, 1994, as cited in Gotlib et al., 1996). Such arguments derive from studies of the differences between anxious and depressed populations on Stroop and memory task performance. A common result is that anxious populations exhibit strong anxiety-related Stroop effects and inconsistent memory biases, whereas depressed populations exhibit inconsistent depression-related Stroop effects, and strong memory

biases (e.g., Dalgleish & Watts, 1990; Mathews & MacLeod, 1994; Mogg & Bradley, 2005; Mineka, Watson, & Clark, 1998).

Qualitative reviews of Stroop performance confirm inconsistent reports of depression-related Stroop effects (Gotlib et al., 1996; Mogg & Bradley, 2005; Williams et al., 1996). In light of the debate around whether depression is associated with an attentional bias, and the equivocal results found for Stroop task performance, it has been suggested that if found, depression-related Stroop effects are more likely to occur under certain conditions: a) when negative semantic primes are presented prior to stimulus presentation in order to activate negative schemas (see Mogg & Bradley, 2005; Segal, 1996; Segal & Swallow, 1994), b) when stimuli are presented for longer periods of time to allow for greater elaborative processing (see Mogg & Bradley, 2005) and/or c) when self-relevant as opposed to general emotional stimuli are presented (see Dalgleish & Watts, 1990; Mathews & MacLeod, 1994; Mogg & Bradley, 2005; Williams et al., 1996).

The literature has converged on predictions about the above aspects of depression-related Stroop performance. However, the impact of population-related variables such as depression severity, age, and gender remains questionable due to a lack of consensus and/or evidence. First, the relationship between depression severity and Stroop performance is unclear given inconsistent findings regarding whether state or trait emotion elicits greater Stroop effects (Gotlib et al., 1996; Williams et al., 1996). Second, age-related differences on the Emotional Stroop task have not been systematically assessed, although MacLeod (1991) determined that interference effects on the Classic Stroop task begin in early childhood, peak around the age of seven or eight, decline throughout adulthood, and begin to increase again after the age of 60. Third, although

there is no evidence in the literature to indicate gender differences on the Classic Stroop task (MacLeod, 1991), women are disproportionately represented in the depression literature, and thus it is worthwhile to examine whether gender moderates effect size on the Emotional Stroop task.

Meta-Analysis as a Statistical Tool

Research results are known to be inconsistent (Schwarzer, 1991), at least partly due to variability in methodologies across studies. Meta-analysis is a statistical technique which aggregates the summary statistics from a number of studies, to draw overall conclusions on the data from a broad literature. The statistical value reported in meta-analysis is the effect size (ES), defined as the standardized mean difference between a criterion group and a comparison group on an outcome variable (Schwarzer, 1991), and divided by an estimate of sample variability (either the comparison group's standard deviation, or a pooled estimate of population deviation). A combined effect size, computed as the average of effect sizes from a series of studies, provides an estimate of the most representative relationship between the groups being compared (Schwarzer, 1991). Meta-analysis takes into account the sample size and the magnitude of the effect size for the comparisons in each study (Rosenthal, 1998). While narrative reviews are potentially subjective and inefficient (Schwarzer, 1991), meta-analysis provides an estimate of overall effects, and of the reliability of these effects (Dobson & Dozois, 2004).

Predictions

Given the mixed pattern of results demonstrated using the Stroop task and debate in the literature, the objective of the current meta-analysis was to address the following

predictions and queries, assuming available data: 1) It was predicted that overall, weak effect sizes would be obtained for depression-related Stroop effects, given the inconsistencies in the literature. 2) According to Beck's and Bower's models, greater depression severity is predicted to result in larger Stroop effects, due to greater accessibility of negative schemas. Despite uncertainty in the empirical literature, it was predicted that greater Stroop effects would be obtained for clinically depressed groups than dysphoric or mood induced groups. 3) Based on Beck's and Bower's models, it was predicted that depressed groups would demonstrate greater Stroop effects for mood-congruent stimuli, than controls. 4) As suggested in the literature, it was predicted that Stroop effects would be greater among depressed groups for primed, self-relevant, and depression-specific words. 5) As suggested in the literature it was predicted that longer stimulus presentations would result in greater depression-related effects than shorter stimulus presentations. 6) What is the impact of gender on Stroop performance? 7) Does age affect Emotional Stroop performance?

Method

This section describes the procedure used to collect articles, to code them for analysis, and the analytic procedures themselves. Each section is presented in turn.

Literature Search

Studies were collected through a search of the online bibliographic search engines PsycINFO and PubMed, using the keywords: 'depression', 'Stroop', 'attention', 'information processing', 'attentional bias', and 'cognitive interference'. The search was restricted to English language articles, published up to and including June 30th, 2007. Unpublished doctoral dissertations were collected in addition to published journal

articles. Using a process of tracking back articles, relevant citations from obtained studies and review articles were also pursued in order to ensure a comprehensive search of the literature base.

Inclusion Criteria

The following inclusion criteria were applied in the selection of studies:

1. The study employed a Classic and/or Emotional Stroop task. Some Emotional Stroop studies included incomparable, variable stimuli, unrelated to depression (e.g., anger, mania, anxiety). In addition, some Classic Stroop studies did not include incongruent colour stimuli, without which it is not possible to make statements about interference. Thus, Emotional Stroop studies were only included if they incorporated stimuli with negative valence and Classic Stroop studies were only included if they incorporated incongruent colour stimuli. Given the wealth of literature indicating the reliability of effects obtained by the Classic Stroop task (MacLeod, 1991), it was included as a benchmark for Emotional Stroop performance.

2. The study contained participants who could be categorized into one of the following groups: clinically depressed, dysphoric, or negatively mood induced. Evidence has not provided support for differences between remitted depressed individuals and non-depressed controls on measures of depressive cognition (Segal & Ingram, 1994), and so studies with remitted depressives were not included. Several studies included mixed depressed samples with a variety of diagnoses of depression, thus it was not possible to analyze Stroop performance by specific depression diagnosis. Depressed samples with comorbid diagnoses of anxiety were included given the high rate of comorbidity between depression and anxiety, and the chance that many depressed participants would have

comorbid anxiety, whether or not it was assessed. Depressed samples with other psychiatric diagnoses (e.g., Schizophrenia) were excluded. Samples with primary health concerns were included if depression was a primary comorbid diagnosis.

3. The study included participants from any age group, and employed any stimulus presentation format, e.g.: card or computer administration, supraliminal unmasked or subliminal masked presentation, word and pictorial stimuli, and primed and unprimed stimuli.

4. The study included a participant control group for between-groups comparisons and/or a stimulus control group for within-groups comparisons.

5. The study included Stroop task performance data prior to treatment, if a treatment was administered as part of the study, because post-treatment comparisons of Stroop performance could be confounded by differing treatment outcomes across studies.

6. The study reported response latency data including sample size of the comparison groups and either means and standard deviations, *t*-test, *F*-test, or exact significance (*p*-values).

7. The study contained participant groups and/or stimuli that met the above criteria and were comparable with at least two other participant groups or sets of stimuli, because a minimum of three comparisons were required to calculate an effect size.

Data Collection and Coding

The following pieces of information were systematically recorded for each study, to the extent that the study provided this information: type of Stroop task (Emotional, Classic), composition of participant groups (number of participants, gender distribution, age), nature of the stimuli (valence, disorder relevance, personal relevance, if primed and

nature of the primes), stimuli presentation procedures (length of stimulus exposure, study design), response latency data, depression scores (Beck Depression Inventory or Hamilton Depression Rating Scale), proportion of women per comparison, and publication status. The proportion of women per comparison was calculated by adding the number of women in the two groups being compared, and dividing this number by the addition of the total number of participants in each group. Study design refers to whether the stimuli were presented randomly, such as in computer presentations where response latency is measured in milliseconds per stimulus, or in a block design, such as in card presentations where response latency is measured in seconds as the time taken to name all of the colours in each condition. Depression scores, the proportion of women per comparison, and length of stimulus exposure were coded as continuous moderator variables. Age, stimuli relevance, publication status, length of stimulus exposure, and study design were coded as categorical moderator variables. Table 1 provides the coding scheme of multi-level variables.

Length of stimulus exposure was recorded in milliseconds and coded as a continuous variable for all studies, in order to account for variable lengths of exposure time. The majority of studies allowed participants as much time as needed to respond to the stimuli. Many of these studies then trimmed the response time data to exclude or truncate response times longer than a specified cut-off, under the assumption that longer response latencies would not be meaningful and that the specified exposure time was more than a sufficient amount of time to respond. The lowest cut-off for erroneous response times was 2000 ms, and thus this number was chosen as the default exposure time for those studies that did not specify length of stimulus exposure. Exposure was also

coded as a categorical variable for those studies that explicitly included supraliminal and subliminal exposure conditions. Studies that did not provide moderator data were excluded from those moderator analyses that required the missing data.

Analyses

Comprehensive Meta-Analysis software, Version 2.002 (CMA; Biostat, Englewood, NJ) was used to perform all analyses. While estimates of ES such as Cohen's (1977) *d* provide a measure of the deviance of the criterion group from the comparison group (Dobson & Dozois, 2004), ESs that employ a weighted estimate of the population standard deviation are more highly recommended, due to their use of a more stable estimate of population variability (Kazdin & Bass, 1989). Hedges *g* is one such form of ES computation, and was utilized in the present study. The cutoff criteria proposed by Cohen (1977) for the identification of small, medium, and large ES are 0.20, 0.50, and 0.80, respectively, and these cutoffs can be applied to Hedges *g*. The 95% confidence intervals for each ES were reported to indicate the range within which the ES would fall if repeatedly calculated. A random effects model was used to represent the data as a range of true effect sizes. A random effects model assumes that the data were drawn from populations that differ from each other in ways that could affect ES, and thus accounts for both within-study error and for true between-study differences (Borenstein, 2005). Under a fixed effects model, one true common effect size is estimated and is therefore less well suited to variable studies (Borenstein, 2005).

Two sets of primary analyses were conducted. The between-groups analyses consisted of comparisons between the depressed and control groups on negative, positive, neutral, and Classic Stroop (incongruent colour) stimuli. In order to minimize the number

of comparisons within a study and the associated ES bias, in the case of a study with multiple depressed and control groups, each depressed group was only compared with its matched control. The within-groups analyses examined differences among the types of Stroop stimuli, within each participant group.

Several secondary analyses were also conducted. A fail-safe *n* was computed for each ES to estimate the number of non-significant results that would be required to render a significant result non-significant (Schwarzer, 1991). Funnel plot analyses were conducted to provide a visual sense of the relationship between effect size and precision (i.e., study size; Borenstein, 2005) for the purpose of detecting the potential influence of publication bias. A funnel plot that indicates the possibility of publication bias would show a thicker clustering of studies on one side of the mean than the other, at the bottom of the graph, reflecting the fact that smaller studies are more likely to be published if they have larger than average effects (Borenstein, 2005). The *Q*-statistic was calculated to determine the dispersion among ESs, as a result of true differences in ES among studies (Borenstein, 2005; Schwarzer, 1991). I^2 was calculated to determine the magnitude of dispersion among ES for each finding. I^2 values of 25, 50, and 75 are considered to represent low, moderate, and high dispersion, respectively (Borenstein, 2005).

Moderator analyses were conducted to determine if differences in study and participant characteristics systematically influenced ES, for those comparisons that were found to have a statistically significant ES, and a statistically significant *Q*. An analysis of variance procedure was employed for the categorical moderator variables, using a fully random effects analysis. Specifically, a random effects model was used to combine studies within each subgroup as well as to combine subgroups to yield the overall effect.

As such, the study-to-study variance was assumed to be the same for all subgroups and this value was computed within subgroups and then pooled across subgroups. The CMA analysis of variance procedure groups the data according to the selected categorical moderator variable and provides an estimate of ES for each group, as well as a *Q'*-value and a *p*-value to indicate whether the difference between groups was significant. A meta-regression procedure, using the unrestricted maximum likelihood model (M. Borenstein, personal communication, August 11, 2008) was employed for the continuous moderator variables. For this analysis, CMA provides a *z*-value and a *p*-value to indicate the nature and significance of the relationship between ES and the moderator variable (i.e., slope). The ES and associated *p*-value for these analyses was based on a smaller subset of analyses from the primary comparisons, because studies that did not report the required data were excluded.

Results

This section describes the dataset that was used for the meta-analysis, the results of the between- and within-groups comparisons with associated secondary analyses, and the results of the moderator analyses. Each section is presented in turn.

Description of the Data

A search of PsycInfo using the previously described keywords yielded 7,025 hits, and a search of PubMed yielded 5,381 hits. There was significant overlap across search results. Thus, the most relevant and least overlapping combinations of keywords were further explored both in PsycInfo and PubMed: 'depression + Stroop', 'depression + cognitive interference', and 'depression + attentional bias'. This more refined search yielded a total of 615 abstracts and from those abstracts, 131 articles which potentially

provided empirical evidence were obtained for review. The final database consisted of 38 studies that provided the data necessary to conduct the meta-analysis (see Appendix). Of these studies, 24% were dissertations, 26% exclusively examined the Classic Stroop task, 53% exclusively examined the Emotional Stroop task, and 21% examined both paradigms. All of the studies included in the analyses involved adult or older adult samples because only one child or adolescent study was found to meet the inclusion criteria for the analyses (Neshat-Doost, Taghavi, Moradi, Yule, & Dalgleish, 1997). In addition, no priming studies were included in the analyses, as there were too few comparable studies to compute an ES. Studies were excluded from the analyses due to: reporting unusable data such as number of correct responses or interference scores, lacking critical information (e.g., number of participants, standard deviations, or exact *p*-values), lacking comparable studies, failing to report Stroop data, reporting changes in data from pre- to post-treatment only, reporting ANOVA interaction terms without follow-up comparisons, lacking a participant or stimulus control group, lacking procedural information rendering data un-interpretable, performing elaborate transformations rendering the data unusable, and collapsing data across participant groups.

Between-Groups Analyses

Each of the three.depressed groups (clinically depressed, dysphoric, sad mood induction) was compared with a control group on four types of Stroop stimuli (negative, positive, neutral and colour incongruent). A comparison between depression-specific and general negative words did not yield significantly different ESs for any of the participant groups, and so the data for both were combined for the remainder of the analyses. Table 2

provides the number of studies (*k*), number of participants (*N*), combined ES (*g*), 95% confidence intervals (*CI*), *Q*, I^2, and fail-safe *n*, per between-groups comparison. The *N* per comparison is inflated in some instances due to multiple comparisons within some studies. There were an insufficient number of studies to calculate a combined ES for sad mood induced versus control participants, on the Classic Stroop.

The comparison of clinically depressed and control participants yielded significant moderate to large ESs across Stroop stimuli. These data were accompanied by significant *Q*'s, large I^2's, and large fail-safe *n*'s. Results for the comparisons between dysphoric and control participants were similarly significant across stimuli, with the exception of the Classic Stroop. The ES's for the comparisons between dysphoric and control participants were slightly smaller (small to moderate) than for the clinically depressed versus control participants, yielded smaller fail-safe *n*'s, and only one significant *Q*, accompanied by a moderate I^2 (for negative stimuli). Only the ES for negative stimuli was significant for the comparisons between sad mood induced and control participants, and was accompanied by a small fail-safe *n* and a non-significant *Q*. Figure 1 presents a graphic depiction of the effect sizes for the between-groups findings.

Within-Groups Analyses

Comparisons were made between the different types of stimuli for each of the three depressed groups and the control group. Table 3 provides the summary information for each within-groups comparison. As with the between-groups comparisons, it should be noted that the *N* per comparison is inflated in some instances due to multiple comparisons within some studies. There was an insufficient number of studies to

calculate a combined ES for dysphoric participants on positive versus neutral stimuli, and for sad mood induced participants on colour incongruent versus control stimuli.

All comparisons between stimuli were significant for clinically depressed participants, except for the comparison between positive and neutral stimuli. While most ES's within this group were small and were accompanied by non-significant *Q*-values and moderate fail-safe *n*'s, the incongruent versus control comparison revealed a large ES, a significant *Q*, a large I^2, and a large fail-safe *n*. The comparisons between colour incongruent and control stimuli revealed the only significant ESs across the other participant groups, each also accompanied by a significant *Q*, a large I^2, and a large fail-safe *n*.

Each between or within-groups comparison was independently entered into a funnel plot analysis, to examine the issue of possible publication bias. These analyses did not reveal a pattern typical of publication bias, for either type of comparison. Nonetheless, some plots indicated that larger studies found ESs below the mean, whereas smaller studies found ESs above the mean. Such a pattern would serve to bias the findings towards smaller ESs. In addition, there was not a large amount of horizontal or vertical scatter, meaning that most studies clustered close to the mean and the distribution of studies was mostly in the moderate to large study size range.

Moderator Analyses

Moderator analyses were performed for five between-groups and three within-groups comparisons that revealed both a significant combined ES and a significant *Q*-value. Age was not found to be a significant moderator for any of the comparisons. The moderating effect of stimuli relevance could not be calculated for seven of the eight

comparisons, because all stimuli were found to be non-self-relevant. Stimuli relevance was not found to be a significant moderator for the one comparison in which there were both self-relevant and non-self-relevant stimuli. The moderating effect of publication status was not calculated for one comparison because all of the studies involved in the comparison were published, and for the remainder of the comparisons publication status was not significant.

Study design revealed significant moderating effects on four of the eight ESs. Specifically, study design significantly moderated all of the comparisons between clinically depressed and control participants. First, study design revealed a significant effect for negative stimuli, Q' (2) = 16.66, $p < 0.001$, where studies with a block design yielded larger effects, $g = 1.42$ ($p < 0.001$), than studies with a random design, $g = 0.75$ ($p < 0.001$), and studies for which the study design was unknown, $g = 0.39$ ($p > 0.05$). Second, study design revealed a significant effect for positive stimuli, Q' (2) = 18.27, $p < 0.001$, where studies with a block design yielded larger effects, $g = 1.36$ ($p < 0.001$), than studies with a random design, $g = 0.41$ ($p > 0.05$), and studies for which the study design was unknown, $g = 0.11$ ($p > 0.05$). Third, study design revealed a significant effect for neutral stimuli, Q' (2) = 9.33, $p < 0.01$, where studies with a block design yielded larger effects, $g = 1.02$ ($p < 0.001$), than studies with a random design, $g = 0.50$ ($p < 0.001$), and studies for which the study design was unknown, $g = 0.22$ ($p < 0.05$). Last, study design revealed a significant effect for colour incongruent (Classic Stroop) stimuli, Q' (2) = 9.52, $p < 0.01$, where studies with a block design yielded smaller effects, $g = 0.23$ ($p > 0.05$), than studies with a random design, $g = 0.72$ ($p < 0.01$), and studies for which the study design was unknown, $g = 0.93$ ($p < 0.001$). None of the studies included in the

comparison between incongruent and control stimuli specified study design for dysphoric participants. The remainder of the comparisons did not reveal a significant moderating effect of study design.

Too few of the studies provided Hamilton Depression Rating Scale (HAM-D) data (21%) for analysis, and so only Beck Depression Inventory (BDI) data were analyzed to determine the moderating effects of depression scores on Stroop performance. BDI, however, only revealed one significant moderating effect. Specifically, it was found that as BDI scores increased, g decreased, $z = 3.76$, $p < 0.001$, for the comparison between clinically depressed and control participants on neutral stimuli ($g = 0.45$, $p < 0.001$). It is not possible to conduct meta-regression with less than three comparisons per analysis, and thus it was not possible to determine the moderating effects of BDI scores for one of the eight comparisons, and the other results were non-significant. Moderator analyses for length of stimulus exposure were calculated for all comparisons to ensure that non-significant findings were not due to imperceptible stimulus exposure. As a continuous moderator, length of stimulus exposure was only found to be a significant predictor for the comparison between clinically depressed and control participants on positive stimuli, $z = 2.26$, $p < 0.05$, and indicated that as length of exposure increased, g also increased ($g = 0.84$, $p < 0.001$). For the remaining comparisons, there was either no variation in exposure times, or exposure was found to be non-significant. When treated as a categorical moderator, exposure was not found to significantly moderate ES either.

The proportion of women per comparison (gender) was found to be a significant moderator for three of the eight ES: a) as proportion of women increased, g decreased, z

= 3.30, $p < 0.001$ for the comparison between clinically depressed and control participants on neutral stimuli ($g = 0.60$, $p < 0.001$), b) as proportion of women increased, g increased, $z = 2.42$, $p < 0.05$ for the comparison between dysphoric and control participants on negative stimuli ($g = 0.46$, $p > 0.05$), and c) as proportion of women increased, g decreased, $z = 5.24$, $p < 0.001$, for the comparison between incongruent and control stimuli for control participants ($g = 1.64$, $p < 0.001$). There were an insufficient number of studies to conduct meta-regression for one of the comparisons and gender was found to be non-significant for the remaining comparisons.

Discussion

The results of the meta-analysis demonstrated large effect sizes for the Emotional Stroop task and medium effect sizes for the Classic Stroop task, when clinically depressed individuals were compared to control groups. Between-groups analyses were affected by both depression severity and stimuli relevance: Stroop performance for the comparisons between clinically depressed individuals and controls resulted in the largest effect sizes, followed by small to medium ESs between dysphoric individuals and controls, and small ESs for those between sad mood induced individuals and controls. The largest between-groups effect sizes were found for negative stimuli, followed by positive, neutral, and Classic Stroop stimuli across comparisons.

The assessment of the robustness of the above results followed the same patterns for depression severity and stimuli relevance, as the largest fail-safe numbers were found for the comparisons between clinically depressed individuals versus controls on negative stimuli, followed by positive, neutral, and Classic Stroop stimuli. Specifically, the number of studies with null effects required to render the effect size for the comparison

insignificant was more than 13 times greater than Rosenthal's (1991) tolerance level for robustness, $5k + 10$ (k = the number of studies included) for negative stimuli, more than 6 times greater for positive stimuli, 4 times greater for neutral stimuli, and 1.5 times greater for Classic Stroop stimuli. The robustness of the effect sizes for the comparisons that involved dysphoric individuals followed the same pattern, although with smaller numbers.

The within-groups analyses demonstrated the largest effect sizes for the comparisons between incongruent and control stimuli (Classic Stroop), across groups. Indeed, these effect sizes were very large, highly significant, and highly robust (all comparisons exceeded Rosenthal's tolerance level). The only other significant effect sizes were for the comparisons between negative and neutral stimuli, and negative and positive stimuli, among clinically depressed participants. These effect sizes were small and highly significant, however, with moderate fail-safe numbers, and these did not meet Rosenthal's (1991) tolerance level for robustness. The moderator analyses were largely non-significant, and for those analyses that revealed significant effects, there was not a consistent pattern of findings.

The number of significant effects demonstrated by this meta-analysis appears to settle the debate as to whether there is a depression-related Emotional Stroop effect, and the magnitude of these effects contradicts the prediction of weak effect sizes. Given the hierarchical nature of the effect sizes obtained for both the within- and between-groups analyses, it appears that the Emotional Stroop distinguishes among levels of depressive experience. As predicted, greater severity is associated with larger effect sizes. Finally,

the meta-analysis appears to demonstrate that the Classic Stroop serves as a good benchmark for comparison with Emotional Stroop performance.

Despite the above overall pattern, several discrepancies in the results need to be addressed. Significant between-groups differences, but non-significant within-groups differences, may indicate that depressed participants do not have an attentional bias toward the experimental stimuli (i.e., negative stimuli), but that the control group has a bias away from the experimental stimuli (Bar-Haim et al., 2007). Given the lack of significant effects for the within-groups comparisons for the non-clinically depressed groups, despite moderate between-groups effects, this explanation may account for the obtained results. In addition, examination of the within-groups effect sizes for the controls showed negative results, which indicates a bias towards neutral, and positive stimuli (i.e., away from the negative stimuli). However, the latter effect sizes were small, and non-significant. In addition, the pattern among comparisons for the clinically depressed group suggested an attentional bias toward the experimental stimuli, as exhibited by significant positive effects between the negative and neutral stimuli, and between the negative and positive stimuli.

An alternative explanation for this pattern of results is that there is a general cognitive slowing among non-clinically depressed groups, which affected their performance across stimuli. In further support of this hypothesis, both the depressed and dysphoric groups demonstrated small to medium significant effect sizes, for neutral stimuli. However, it should be noted that there was at least some content-specificity reflected in larger effect sizes for the negative content than the positive or neutral content. The fact that the clinically depressed group demonstrated within-group effects, albeit

small, may indicate a qualitative difference between clinically depressed and dysphoric or mood induced groups, such that only clinically depressed individuals exhibit an attentional bias for emotion-congruent stimuli. This conclusion is compatible with the idea that attentional bias to mood-congruent stimuli is implicated in the etiology and maintenance of depression (Gotlib et al., 1996).

The few moderator analyses that yielded significant results did not demonstrate a consistent pattern. The variable that moderated the greatest number of comparisons was study design. In three of the eight analyses, block design resulted in larger effects than random or unknown designs. However, these larger effects were not consistently significant, nor were the effects associated with random and unknown designs consistently non-significant. Other studies have also found that block design yields a significantly larger effect size than random design (Bar-Haim et al., 2007; Williams et al., 1996). While studies that used block design yielded a larger number for average response time per participant group or stimulus condition, than studies that reported response times for individual stimuli, the difference between groups or stimuli should have remained constant given an equal number of stimuli per card. This finding may be explained by differences in the mechanics of each presentation method, and thus it is something to consider when designing future studies.

Taken together, three findings from the meta-analysis raise questions as to whether it is self-relevance and disorder-congruence that determine which stimuli are favoured on the Emotional Stroop task. First, self-relevance did not significantly moderate any of the comparisons. That said, only four studies included stimuli that were participant generated or rated as self-relevant, thus it is difficult to make conclusive

statements about this finding. Second, no significant differences in effect size were found between depression-specific and general negative stimuli, and there was not a consistent pattern in terms of which of the two yielded larger effect sizes. Finally, the primary analyses revealed large effect sizes for comparisons between depressed and control groups for both negative and positive stimuli, and non-significant or small effect sizes for comparisons between negative and positive stimuli among depressed groups. Contrary to prediction, the latter results revealed that depressed individuals experience an attentional bias for both negative and positive stimuli, which suggests a general emotional bias for depressed individuals on the Stroop, rather than a disorder-specific and self-relevant attentional bias. It should be noted, though, that the negative stimuli used in some studies were characterized as interpersonal (Dozois & Dobson, 2001; Hamilton, 2003), autonomous, or sociotropic (Gupta-Rogers, 1999; Hamilton, 2003; Kinderman, 1994; Nunn, Mathews, & Trower, 1999; Shapiro, 2002), which may have increased the relevance of the stimuli.

The moderator analyses that examined depression severity, age, gender, and length of stimulus exposure did not yield significant results for the most part, and those that did were inconsistent. Many studies did not indicate which version of the Beck Depression Inventory they used and thus it is possible that the depression severity results were affected by the use of different versions of the measure. Age may not have been a significant moderator due to a lack of age variability in the samples. There were insufficient child or adolescent studies to be analyzed, and studies with older adults included samples typically aged 60 or 65 through 85. MacLeod (1991) deduced that performance on the Stroop decreases after the age of 60 and thus with a mixed age

sample, the performance of 'younger' older adults would attenuate any age effects that may be present at older ages. The lack of age effects corroborated evidence for the meta-analysis conducted by Verhaeghen and De Meersman (1998), who did not find any age-related interference effects on the Stroop for older compared with younger adults. Although gender moderated three comparisons, the lack of consistent findings across comparisons precludes a confident statement about the potential moderating effects of gender.

The present meta-analysis found Stroop effects regardless of exposure length, which suggests that the stimuli were processed without the need for conscious perception and more elaborative processing, contrary to prediction. This result corroborates previous research that found that Stroop interference does not depend on conscious strategies (Bar-Haim et al., 2007; Williams et al., 1996). Last, none of the studies included in the meta-analysis incorporated priming procedures, and yet several significant and robust effects were found, once again, contrary to prediction.

Cognitive Mechanisms Underlying Depression-Related Bias on the Stroop Task

Although the Emotional Stroop task is widely used in the investigation of attentional bias, it has been criticized by several researchers (e.g., Algom, Chajut, & Lev, 2004; de Ruiter & Brosschot, 1994; Isaacowitz, 2007; MacLeod, Mathews, & Tata, 1986). The argument is that delayed response latencies observed on the Stroop task may be unrelated to attention or early information-processing processes, but rather to late processes (de Ruiter & Brosschot, 1994; MacLeod, Mathews, & Tata, 1986) such as cognitive avoidance or response inhibition (de Ruiter & Brosschot, 1994). The Parallel Distributed Processing (PDP) model (MacLeod, 1991; Williams et al., 1996), argues that

the Stroop task taps both attentional and response processes, and that Stroop effects should not be assumed to represent an attentional bias. Bar-Haim et al. (2007) presented a similar model for their anxiety-related Stroop findings, which they argued "suggest that strong claims that bias in only one stage of processing accounts for the attentional bias in anxiety should be toned down" (p. 17). Their model posits that Stroop effects may be related to preattentive, attentional, and postattentive processes. The present meta-analysis does not indicate whether the obtained depression-related Stroop effects represent an attentional bias or a response bias to emotional stimuli. In order to accurately parse out the underlying cognitive mechanisms to specifically determine if there is in fact a depression-related attentional bias, it would be necessary to examine depression-related results obtained from a 'pure' measure of attentional bias, as discussed below.

Study Strengths and Limitations

The present meta-analysis has several significant strengths. To start, it is the first quantitative review of the depression-related Stroop literature. Previous qualitative reviews came to the conclusion that the depression-related Stroop literature was inconsistent, thus indicating the need to conduct an objective quantitative review such as the present study. The results of the meta-analysis provide strong evidence on the state of the science, and are able to speak to many questions in the field. The scope of the literature review was sufficiently broad to represent the discipline, and the stringent inclusion criteria for the analyses ensured quality of the data. Data collection and entry was carefully conducted, and more conservative analyses were selected, thereby ensuring confidence in the results.

Despite the above strengths, potential limitations of the present meta-analysis need discussion. Meta-analysis as a statistical tool has been criticized on several grounds. First, it has been argued that meta-analysis exacerbates the "file drawer problem" due to its frequent reliance on published research. This problem may lead to an overestimation of effect size, due to an over-representation of published studies, which more often have significant effects than unpublished studies (Schwarzer, 1991). The present meta-analysis utilized three methods to address this issue: unpublished dissertations were included, funnel plot analyses were conducted, and the potential moderation effects of publication status were assessed. The results of these methods and analyses indicate that the 'file drawer problem' did not significantly bias the findings.

Second, because meta-analysis involves the combination of a large pool of data from varied research, an effect size may be misleading because some of the contributing data may be faulty or weak (Schwarzer, 1991). Similarly, due to the scope required to derive meaningful conclusions, comparisons may be made among studies that differ appreciably in design and method (Schwarzer, 1991). The inclusion and exclusion criteria and the moderator analyses employed in the present meta-analysis represented an effort to balance this potential bias. Most of the significant effect sizes were tempered by significant Q-values and high I^2-values. These numbers reveal that there is a significant amount of true variation in effect sizes between studies, which cannot be accounted for by random error. Previous qualitative reviews have commented on the inconsistent findings in the field (Gotlib et al., 1996; Mogg & Bradley, 2005; Williams et al., 1996). The results of the current study reflect the methodological variability in the literature base. Studies varied as to whether and how they transformed their data, in terms of what

stimuli they used (which differed in content, number, colour, match of conditions, presentation method, and response format), and in terms of their samples (which differed on variables such as severity of depression, diagnosis, level of symptoms, mood induction procedure, age, and gender).

The control groups used in the available research likely exerted an effect on effect size estimates (Bar-Haim et al., 2007). Specifically, clinical groups were often matched with non-clinical or 'healthy' controls. In contrast, dysphoric individuals were typically matched with comparable but non-dysphoric individuals who did not meet the cut-off criteria on the same measure. Sad mood induced participants were typically matched with similar non-mood induced or neutral mood induced counterparts. The magnitude of any effect size is partially determined by the degree of difference between the two samples employed in the study. Thus, the differences in effect sizes observed here may be due in part to sample differences.

The most common criticism of meta-analysis is that it amalgamates non-independent data, and thereby introduces bias into effect size calculations (Schwarzer, 1991). Non-independent data derive from at least three sources: multiple outcomes from the same studies, comparisons among more than one criterion group and a single control group (Schwarzer, 1991), and multiple publications by the same author with parts of the same data set. The amalgamation of non-independent data inflates effect sizes because sample size is artificially increased in these studies, which are then assigned more weight. Moreover, when multiple studies by the same author are included, potential biases of the author may be transferred into the analyses, through their data, which tend to support the author's theoretical stance (i.e., researcher allegiance). Although the problem of non-

independence may be negligible (Schwarzer, 1991), the issue of non-independence was addressed here by the use of more conservative statistics (i.e., Hedge's *g*, random effects and fully random effects models). Furthermore, the number of comparisons per study was reduced by only including matched criterion and control group pairs.

Future Directions

The current meta-analysis provides sufficiently large and robust effect sizes to argue that if any further research on the Stroop task is pursued, it need not address questions regarding Stroop performance with emotional stimuli, among clinically depressed populations. The utility of further examining Stroop performance among dysphoric or mood induced populations is also questionable, given moderate effect sizes. If there is interest in further research on the Stroop task in depression, it should address remaining gaps in the literature base. For example, the depression-related Stroop performance in child and adolescent populations remains unclear. In addition, future research could focus on the variables that moderate depression-related Stroop performance. In particular, while depression and anxiety are highly comorbid, few studies have compared an anxious group with a depressed group on Stroop performance. Moreover, the variability in anxiety measures used in Emotional Stroop research to date precludes the assessment of self-reported anxiety as a potential moderating variable of effect size.

However, if the field's interest is the examination of attentional bias rather than Stroop performance *per se*, the dot probe task and eye tracking methods may provide a more accurate assessment of attentional biases than the Stroop task. In the dot probe task, two words are simultaneously presented at two different locations on a visual display

(MacLeod, 1986). Immediately following stimulus presentation, a visual probe replaces one of the words. Detection latency is presumed to vary as a function of stimuli relevance and location, and indicates whether visual attention was directed to or away from the stimuli of interest. The dot probe task has received support in the literature as a more accurate measure of attentional bias than the Stroop task (Bar-Haim et al., 2007; de Ruiter & Brosschot, 1994; Mathews & MacLeod, 1994; Mineka & Sutton, 1992; Mineka, Watson, & Clark, 1998).

Due to concerns that the dot probe task is also an indirect measure of attentional bias (Eizenman et al., 2003; Isaacowitz, 2007), Eizenman et al. (2003) developed an eye-tracking technology to continuously monitor point-of-gaze. They postulated that attentional bias among depressed populations would be typified by more time spent visually fixating on negative stimuli, and by difficulties in shifting visual gaze away from such stimuli, as compared with normal controls. They found that although depressed individuals spent significantly more time gazing at dysphoric stimuli, they did not scan dysphoric stimuli more frequently than normal controls. Eizenman et al. (2003) interpreted these findings to mean that depression primarily influences later stages of processing, as opposed to early attentional processes. While the method and associated findings of this study are compelling, more research is required before conclusive statements can be made about the paradigm, or the implications of the resulting data.

Conclusion

The present meta-analysis demonstrated that there are large and robust depression-related Stroop effects. Although these effects do not reflect an emotion-congruent bias, they do distinguish among levels of depressive experience, where greater

severity is associated with larger effect sizes. Moreover, these effects do not require priming procedures, longer stimulus exposure, or the presentation of self-relevant or disorder-congruent stimuli, to be obtained. Heterogeneity within the findings could not be fully accounted for by the moderating effects of gender, age, publication status, study design, or self-reported depression. This finding reflects both inconsistent measurement of these moderators, and heterogeneity among results in the original studies. The meta-analysis also demonstrated that the Classic Stroop serves as a good benchmark for comparison with Emotional Stroop performance. Given these findings, further research on the Stroop task in depression is not necessary. Research using a more direct measure of depression-related attentional bias is recommended.

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General, 133*, 323-338.

Andreotti, P. A. (2000). Effects of angry mood on attention and recall. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(6-B), 3268. (UMI No. 9977086)

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1-24.

Baune, B. T., Suslow, T., Engelien, A., Arolt, V., & Berger, K. (2006). The association between depressive mood and cognitive performance in an elderly general population – the MEMO study. *Dementia and Geriatric Cognitive Disorders, 22*, 142-149.

Beck, A. T., Rush, J. A., Shaw, B. F., & Emery, G. (1979). Preface. In A. T. Beck, J. A. Rush, B. F. Shaw, & G. Emery (Eds.), *Cognitive therapy of depression* (pp. 6-33). New York, NY: The Guilford Press.

Boissevain, M. D. (1995). Information processing in chronic pain: The role of depression. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*(1-B), 0517. (UMI No. 0622)

Boone, K. B., Lesser, I. M., Miller, B. L., Wohl, M., Berman, N., Lee, A., et al. (1994). Cognitive functioning in a mildly to moderately depressed geriatric sample:

Relationship to chronological age. *Journal of Neuropsychiatry & Clinical Neurosciences, 6*, 267-272.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-analysis Version 2.002* (Manual and Computer Software). Englewood, NJ: Biostat.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129-148.

Bradley, B. P., Mogg, K., Millar, N., & White, J. (1995). Selective processing of negative information: Effects of clinical anxiety, concurrent depression, and awareness. *Journal of Abnormal Psychology, 104*(3), 532-536.

Broomfield, N. M., Davies, R., MacMahon, K., Ali, F., & Cross, S. M. B. (2007). Further evidence of attention bias for negative information in late life depression. *International Journal of Geriatric Psychiatry, 22*, 175-180.

Clark, D. A., Beck, A. T., & Alford, B. (1999). Scientific foundations of cognitive theory and therapy of depression. Hoboken, NJ: John Wiley & Sons, Inc.

Cohen, J. D. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Bulletin, 97*, 332-361.

Constant, E. L., Adam, S., Seron, X., Bruyer, R., Seghers, A., & Daumerie, C. (2006). Hypothyroidism and major depression: A common executive dysfunction. *Journal of Clinical and Experimental Neuropsychology, 28*(5), 790-807.

Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The addiction-Stroop test: Theoretical considerations and procedural recommendations. *Psychological Bulletin, 132*, 443-476.

Dalgleish, T., & Watts, F. N. (1990). Biases of attention and memory in disorders of anxiety and depression. *Clinical Psychology Review, 10*, 589-604.

de Ruiter, C., & Brosschot, J. F. (1994). The emotional stroop interference effect in anxiety: Attentional bias or cognitive avoidance? *Behaviour Research and Therapy, 32*, 315-319.

den Hartog, H. M. D., Derix, M. A., van Bemmel, A. L., Kremer, B., & Jolles, J. (2003). Cognitive functioning in young and middle-aged unmedicated out-patients with major depression: Testing the effort and cognitive speed hypotheses. *Psychological Medicine, 33*, 1443-1451.

Dieckmann, L. L. (1991). Inhibition of neutral and emotional stimuli in depression and aging. *Dissertation Abstracts International, 52*(5-B), 2771.

Dijkstra, J. B., Strik, J. J. M. H., Lousberg, R., Prickaerts, J., Riedel, J., Jolles, J., et al. (2002). Atypical cognitive profile in patients with depression after myocardial infarction. *Journal of Affective Disorders, 70*, 181-190.

Dobson, K. S., & Dozois, D. J. A. (2004). Attentional biases in eating disorders: A meta-analytic review of Stroop performance. *Clinical Psychology Review, 23*, 1001-1022.

Dozois, D. J. A. (1999). Cognitive organization and information processing in clinical depression: The structure and function of sociotropic schemata (Doctoral dissertation, University of Calgary, 1999). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(4-B), 2232.

Dozois, D. J. A., & Dobson, K. S. (2001). Information processing and cognitive organization in unipolar depression: Specificity and comorbidity issues. *Journal of Abnormal Psychology, 110*(2), 236-246.

Dudley, R., O'Brien, J., Barnett, N., McGuckin, L., & Britton, P. (2002). Distinguishing depression from dementia in later life: A pilot study employing the emotional Stroop task. *International Journal of Geriatric Psychiatry, 17*(1), 48-53.

Eizenman, M., Yu, L. H., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., et al. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder. *Psychiatry Research, 118*, 117-128.

Feil, D., Razani, J., Boone, K., & Lesser, I. (2003). Apathy and cognitive performance in older adults with depression. *International Journal of Geriatric Psychology, 18*, 479-485.

Gilboa-Schechtman, E., Revelle, W., & Gotlib, I. H. (2000). Stroop interference following mood induction: Emotionality, mood congruence, and concern relevance. *Cognitive Therapy and Research, 24*(5), 491-502.

Gotlib, I. H., & Cane, D. B. (1987). Construct accessibility and clinical depression: A longitudinal investigation. *Journal of Abnormal Psychology, 96*(3), 199-204.

Gotlib, I. H., & McCann, C. D. (1984). Construct accessibility and depression: An examination of cognitive and affective factors. *Journal of Personality and Social Psychology, 47*(2), 427-439.

Gotlib, I. H., Roberts, J. E., & Gilboa, E. (1996). Cognitive interference in depression. In I. G. Sarason, G. R. Pierce, & B. R. Sarason (Eds.), *Cognitive interference: Theories, methods, and findings* (pp. 347-377). Hillsdale, NJ, England: Lawrence Erlbaum Associates Inc.

Gupta-Rogers, R. (1999). The nature, specificity, and temporal stability of emotional information processing in sociotropic and independent women. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*(1-B), 0382. (UMI No. 612354709)

Hamilton, K. E. (2003). Cognitive vulnerability to depression: Accessibility on information processing biases in remitted depression (Doctoral dissertation, University of Calgary, 2003). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*(1-B), 437.

Isaacowitz, D. M. (2007). Understanding individual and age differences in well-being: An experimental attention-based approach. In A. D. Ong, & M. H. M. van Dulmen (Eds), *Oxford handbook of methods in positive psychology* (pp. 220-232). New York, NY: Oxford University Press.

Janer, K. W. (1995). Attentional bias, implicit memory, and general slowing in depression. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 55*(11-B), 5072. (UMI No. 9509917)

Julian, L. J., & Mohr, D. C. (2006). Cognitive predictors of response to treatment for depression in multiple sclerosis. *Journal of Neuropsychiatry and Clinical Neurosciences, 18*, 356-363.

Kalayam, B., & Alexopoulos, G. S. (2003). A preliminary study of left frontal region error negativity and symptom improvement of geriatric depression. *American Journal of Psychiatry, 160*, 2054-2056.

Katz, L. J., Wood, D. S., Goldstein, G., Auchenbach, R. C., & Geckle, M. (1998). The utility of neuropsychological tests in evaluation of attention-deficit/hyperactivity disorder (ADHD) versus depression in adults. *Assessment, 5*, 45-51.

Kazdin, A. E., & Bass, D. (1989). Power to detect differences between alternative treatments in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 138-147.

Kerr, N., Scott, J., & Phillips, M. L. (2005). Patterns of attentional deficits and emotional bias in bipolar and major depressive disorder. *British Journal of Clinical Psychology, 44*, 343-356.

Kinderman, P. (1994). Attentional bias, persecutory delusions and the self-concept. *British Journal of Medical Psychology, 67*, 53-66.

Klieger, D. M. & Cordner, M .D. (1990). The Stroop task as measure of construct accessibility in depression. *Personality and Individual Differences, 11*(1), 19-27.

Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: A review and meta-analysis. *Neuropsychology, 21*, 251-262.

Lemelin, S., Baruch, P., Vincent, A., Everett, J., & Vincent, P. (1997). Distractibility and processing resource deficit in major depression: Evidence for two deficient attentional processing models. *Journal of Nervous and Mental Disease, 185*, 542-548.

Lemelin, S., Baruch, P., Vincent, A., LaPlante, L., Everett, J., & Vincent, P. (1996). Attention disturbance in clinical depression: Deficient distractor inhibition or processing resource deficit? *The Journal of Nervous and Mental Disease, 184*, 114-121.

Lepage, J. P. (1999). A comparison of schema and network models and the processing of affective personal information. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*(1-B), 0370. (UMI No. 9917116)

Lim, S. L., & Kim, J. H. (2005). Cognitive processing of emotional information in depression, panic, and somatoform disorder. *Journal of Abnormal Psychology, 114*, 50-61.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15-20.

Mathews, A., & MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology, 45*, 25-50.

McNeil, D. W., Tucker, P., Miranda, R. Jr., Lewin, M. R., & Nordgren, J. C., (1999). Response to depression and anxiety Stroop stimuli in posttraumatic stress disorder, obsessive-compulsive disorder, and major depressive disorder. *Journal of Nervous and Mental Disease, 187*(8), 512-516.

Mineka, S., & Sutton, S. K. (1992). Cognitive biases and the emotional disorders. *Psychological Science, 3*(1), 65-69.

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology, 49*, 377-412.

Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Therapy and Research, 29*(1), 29-45.

Mogg, K., Bradley, B. P., Williams, R., & Mathews, A. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102*(2), 304-311.

Neshat-Doost, H. T., Taghavi, M. R., Moradi, A. R., Yule, W., & Dalgleish, T. (1997). The performance of clinically depressed children and adolescents on the modified Stroop paradigm. *Personality and Individual Differences, 23*(5), 753-759.

Nunn, J. D., Mathews, A., & Trower, P. (1997). Selective processing of concern-related information in depression. *British Journal of Clinical Psychology, 36*, 489-503.

Pérez, M. G., Rivera, R. M. B., Fuster, A. B., & Rodríguez, M. A. R. (1999). Attentional biases and vulnerability to depression. *The Spanish Journal of Psychology, 2*(1), 11-19.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, California: Sage Publications Inc.

Rosenthal, R. (1998). Meta-analysis: Concepts, corollaries and controversies. In J. Adair & D. Belanger (Eds.), *Advances in psychological science: Social, personal and cultural aspects, Vol. 1* (pp. 371 – 384). Hove, UK: Psychology Press.

Schwarzer, R. (1991). *Meta-Analysis Version 5.3* (Manual). Berlin, Germany: Free University.

Segal, Z. V. (1996). Cognitive interference in depressive and anxiety-based disorders. In I. G. Sarason, G. R. Pierce, & B. R. Sarason (Eds.), *Cognitive interference: Theories, methods, and findings* (pp. 325-345). Hillsdale, NJ, England: Lawrence Erlbaum Associates Inc.

Segal, Z. V., & Ingram, R. E. (1994). Mood priming and construct activation in tests of cognitive vulnerability to unipolar depression. *Clinical Psychology Review, 14*(7), 663-695.

Segal, Z. V., & Swallow, S. R. (1994). Cognitive assessment of unipolar depression: Measuring products, processes and structures. *Behaviour Research and Therapy, 32*, 147-158.

Shapiro, A. M. (2002). Attentional biases for negative information in depression: Rumination and deficient inhibitory processes. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(8-B), 3815. (UMI No. 3021946)

Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology, 52*, 63-76.

Silberman, C. D., Laks, J., Capitão, C. F., Rodrigues, C. S., Moreira, I. Vasconcellos, L. F. R., et al. (2007). Frontal functions in depressed and nondepressed Parkinson's disease patients: Impact of severity stages. *Psychiatry Research, 149*, 285-289.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Szöke, A., Trandafir, A., Dupont, M.-E., Méary, A., Schürhoff, F., & Leboyer, M. (2008). Longitudinal studies of cognition in schizophrenia: Meta-analysis. *The British Journal of Psychiatry, 192*, 248 - 257.

Verhaeghen, P. & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging, 13*, 120-126.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*(1), 3-24.

Yovel, I., & Mineka, S. (2004). Hierarchical models of emotional disorders and emotion-congruent cognitive biases. *Personality and Individual Differences, 36*, 679-694.

Table 1

Coding Scheme for Multi-level Variables

Variable	Coding	Description
Participant groups	Clinically depressed	Inpatients or outpatients with a current primary diagnosis of unipolar depression, or research participants who met Structured Clinical Interview for Diagnosis criteria for current unipolar depression (including major depressive disorder, major depressive episode, dysthymia, minor depression, and major depression).
	Dysphoric	Participants that were assigned to a dysphoric group based on an elevated score on a depression inventory, above a specified threshold.
	Sad Mood	Participants that underwent a negative (sad) mood induction prior to completing the Stroop task.
	Control	All non-specific control populations: labeled as non-psychiatric, non-depressed, healthy controls, less than cut-off score, etc.
	Control Neutral	Participants that underwent a neutral mood induction prior to completing the Stroop task.

	Control Positive	Participants that underwent a positive mood induction prior to completing the Stroop task.
Participant age	Adult	Ages 18 through 60
	Older Adult	Age 60 and older
	Children/Adolescent	Age less than 18
Publication status	Published	Empirical journal article
	Unpublished	Doctoral dissertation
Stimuli relevance	Nonself-relevant	Participants did not select stimuli or rate them as personally relevant
	Self-relevant	Participants selected stimuli themselves or rated them as personally relevant
Study design	Block	Stimuli presented in category blocks and response time recorded per block.
	Random	Stimuli presented individually and randomly, response time recorded per stimulus.
	Unknown	Study design was not specified and/or unclear.

Table 2

Effect Size, Homogeneity, and Publication Bias Results for Between-Groups Comparisons

Comparison	*k*	*N*	*g*	95% CI	*Q* (df)	I^2	Fail-safe *n*
Negative content							
Clinically Depressed vs. Control	17	1056	0.97***	0.73 – 1.20	77.04 (24)***	68.85	1270
Dysphoric vs. Control	6	467	0.55***	0.26 - 0.84	15.71 (8)*	47.08	54
Sad Mood vs. Control	4	445	0.19*	0.02 – 0.37	12.95 (15)^{ns}	0	2
Positive content							
Clinically Depressed vs. Control	13	837	0.84***	0.51 – 1.17	93.76 (18)***	80.80	518
Dysphoric vs. Control	3	185	0.65***	0.36 – 0.95	2.56 (4)^{ns}	0	22
Sad Mood vs. Control	4	475	0.13^{ns}	-0.04 – 0.29	10.80 (15)^{ns}	0	0
Neutral content							
Clinically Depressed vs. Control	15	973	0.55***	0.35 – 0.76	57.75 (23)***	60.17	387

Dysphoric vs. Control	4	320	0.26*	0.02 – 0.51	5.39 (5)^{ns}	7.27	6
Sad Mood vs. Control	2	387	0.15^{ns}	0.03 – 0.34	9.41 (13)^{ns}	0	0
Classic Stroop							
Clinically Depressed vs. Control	9	706	0.53***	0.25 – 0.80	29.97 (10)**	66.63	104
Dysphoric vs. Control	2	421	0.27^{ns}	-0.02 – 0.55	1.62 (2)^{ns}	0	1

Note: *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$, ns = non-significant, k = number of studies, N = number of participants, g = combined effect size, CI = confidence intervals, Q = significance of between-study dispersion, I^2 = magnitude of between-study dispersion, fail-safe n = number of additional studies with null effects required to render the results non-significant.

Table 3

Effect Size, Homogeneity, and Publication Bias Results for Within-Groups Comparisons

Comparison	*k*	*N*	*g*	95% CI	*Q* (df)	I^2	Fail-safe *n*
Clinically Depressed							
Negative vs. Neutral	16	542	0.25***	0.13 – 0.37	20.53 (26)^{ns}	0	67
Negative vs. Positive	15	450	0.21**	0.08 – 0.34	6.19 (20)^{ns}	0	37
Positive vs. Neutral	11	358	0.08^{ns}	-0.06 – 0.22	10.60 (17)^{ns}	0	0
Incongruent vs. Control	12	502	1.67***	1.14 – 2.20	175.33 (14)***	92.02	1656
Dysphoric							
Negative vs. Neutral	4	182	0.16^{ns}	-0.04 – 0.37	0.97 (5)^{ns}	0	0
Negative vs. Positive	3	83	-0.02^{ns}	-0.32 – 0.28	0.69 (4)^{ns}	0	0
Incongruent vs. Control	2	66	2.67***	1.40 – 3.95	11.99 (2)**	83.32	97

Sad Mood Induction							
Negative vs. Neutral	4	284	0.10^{ns}	-0.04 – 0.24	4.45 (16)^{ns}	0	0
Negative vs. Positive	4	289	0.11^{ns}	-0.03 – 0.25	3.85 (16)^{ns}	0	0
Positive vs. Neutral	4	299	-0.05^{ns}	-0.19 – 0.08	1.55 (16)^{ns}	0	0
Control							
Negative vs. Neutral	18	788	-0.01^{ns}	-0.11 – 0.09	9.26 (40)^{ns}	0	0
Negative vs. Positive	16	618	-0.02^{ns}	-0.14 – 0.10	10.23 (24)^{ns}	0	0
Positive vs. Neutral	14	560	0.01^{ns}	-0.11 – 0.14	2.54 (22)^{ns}	0	0
Incongruent vs. Control	12	769	1.89***	1.42 – 2.37	157.55 (13)***	91.75	2793

Note: *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$, ns = non-significant, k = number of studies, N = number of participants, g = combined effect size, CI = confidence intervals, Q = significance of between-study dispersion, I^2 = magnitude of between-study dispersion, fail-safe n = number of additional studies with null effects required to render the results non-significant.

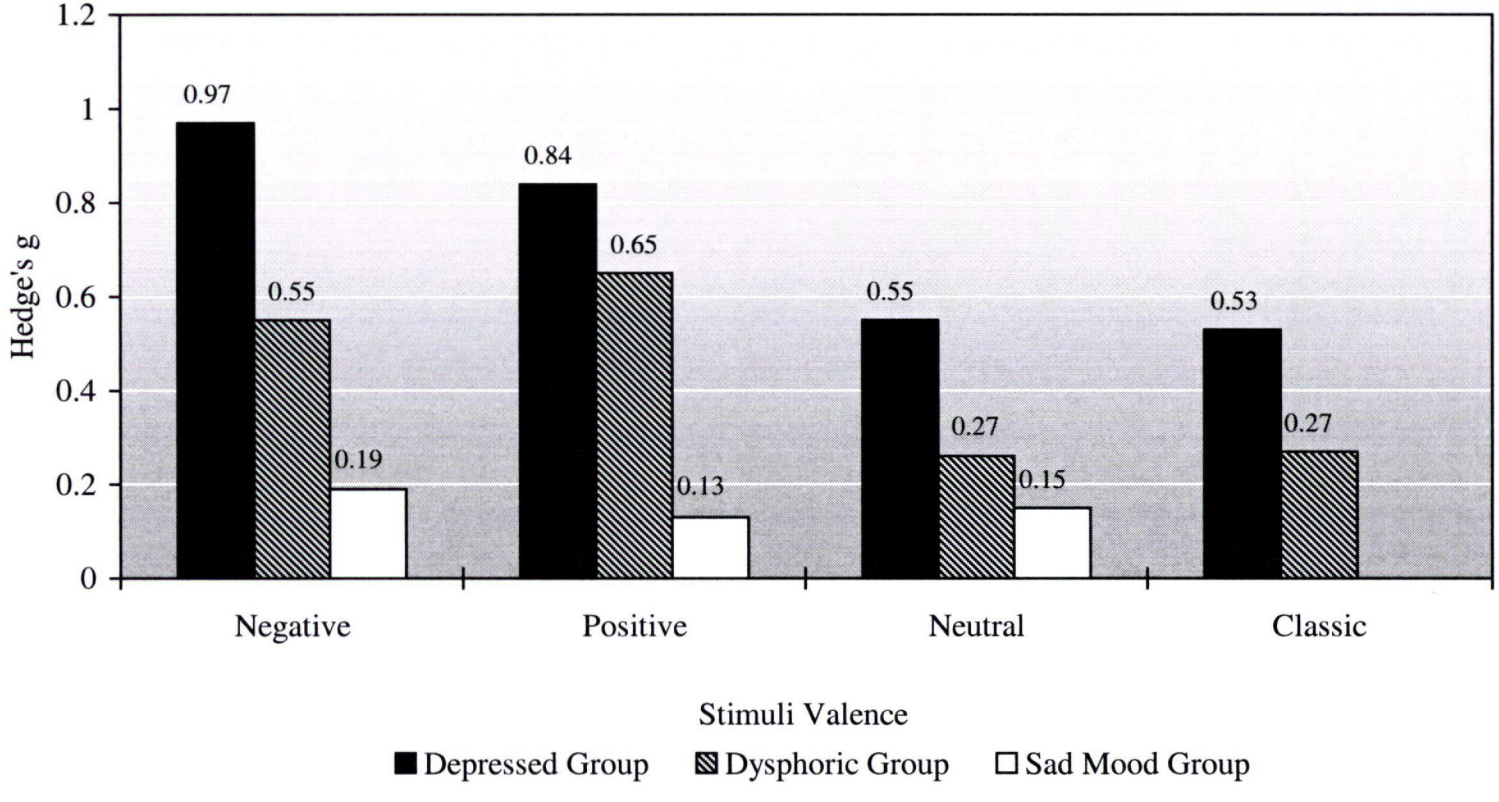

Figure 1. Effect Sizes for Between-Groups Comparisons.

Appendix

Individual Studies Included in the Meta-Analysis

Study	Task	Sample	Participant Groups (sample size)	Stimuli	Presentation
Andreotti (2000)	Emotional	Students	1. Angry Mood Induction (25) 2. Sad Mood Induction (25) 3. Neutral Mood Induction (25)	a) Angry b) Positive c) Neutral d) Non-lexical	Card presentation
Baune et al. (2006)	Classic	Community	1. Dysphoric (36) 2. Not Dysphoric (328)	a) Words b) Colour/word-incongruent	Not specified
Boissevain (1995)	Emotional	Clinical	Principle Study 1. Chronic Pain/Depressed (15) 2. Chronic Pain/Not Depressed (15) 3. No Pain/Depressed (15) 4. No Pain/Not Depressed (15)	Principle Study a) Positive b) Neutral c) Sensory Pain d) Affective Pain e) Depressive	Card presentation
Boone et al. (1994)	Classic	Clinical / Community	1. Age 46-59 a) Major Depressive Disorder (36) b) Controls (58) 2. Age 60-69 a) Major Depressive Disorder (23) b) Controls (54) 3. Age 70-85 a) Major Depressive Disorder (14) b) Controls (41)	a) Words b) Colours c) Colour/word-incongruent	Not specified

Bradley et al. (1995)	Emotional	Clinical	1. GAD (11) 2. GAD + Depressed (principle diagnosis and comorbid MDD or dysthymia) (9) 3. Controls (20)	a) Anxious b) Depressed c) Categorized Neutral d) Uncategorized Neutral	Subliminal & Supraliminal Random design
Broomfield et al. (2007)	Emotional	Clinical / Community	1. Major Depressive Disorder (16) 2. Controls (19)	a) Negative b) Neutral c) Positive	Computer presentation
Constant et al. (2006)	Emotional & Classic	Clinical	1. Remission from thyroid carcinoma treatment (23) 2. Major Depressive Episode (20) 3. Controls (26)	Classic a) Coloured symbols b) Colour/word-congruent c) Colour/word-incongruent Emotional a) Colour/word-congruent b) Neutral c) Depression d) Anxiety	Computer presentation Subliminal & Supraliminal for 3 emotional Stroop stimuli
den Hartog et al. (2003)	Classic	Clinical	1. MDD (30) 2. Healthy Controls (38) 3. Severe Allergic Rhinitis (25)	a) Colour words b) Colour patches c) Colour/word-incongruent	Card presentation
Dieckman (1991)	Emotional	Students / Community	Experiment 2 1. Young – Non-dysphoric (15) 2. Young – Dysphoric (15) 3. Old – Non-dysphoric (16) 4. Old – Dysphoric (9)	Experiment 2 a) Positive Self-descriptive b) Negative Self-descriptive c) Neutral d) Nonlexical	Computer presentation

Dijkstra et al. (2002)	Classic	Clinical	1. Non-depressed Controls (48) 2. Non-depressed – after first mycordial infarction (48) 3. Depressed - after first mycordial infarction (48)	a) Colour words b) Colour patches c) Colour/word-incongruent	Card presentation
Dozois (1999)	Emotional & Classic	Clinical	Experiment 1 1. Major Depressive Disorder (50) 2. Anxious Non-depressed (25) 3. Non-depressed, Non-psychiatric (25) Experiment 2 1. Stable Depressed (22) 2. Remitted Depressed (23)	Experiment 1 & 2 a) Positive b) Negative c) Colour/word-congruent d) Colour/word-incongruent e) Nonlexical	Computer presentation
Dozois & Dobson (2001)	Emotional	Community	1. Current MDD & comorbid anxiety disorder (26) 2. Current MDD & no comorbid anxiety disorder (24) 3. Current anxiety disorder & no MDD (25) 4. Non-psychiatric controls (25)	a) Colour/word-congruent b) Colour/word-incongruent c) Interpersonal positive d) Interpersonal negative e) Non-lexical	Computer presentation
Dudley et al. (2002)	Emotional	Clinical	1. MDD (12) 2. Alzheimer's Disease (12) 3. Controls (12)	a) Coloured x's b) Neutral c) Positive d) Negative	Card presentation

Feil et al. (2003)	Classic	Clinical / Community	1. MDD (89)	a) Words b) Colour patches c) Colour/word-incongruent	Not specified
Gilboa-Schechtman et al. (2000)	Emotional	Students	1. Negative Mood Induction 2. Positive Mood Induction	1. Experimenter-provided: a) neutral b) negative c) positive 2. Participant-generated: a) neutral b) negative c) positive	Computer presentation
Gotlib & Cane (1987)	Emotional	Clinical / Community	1. Major depressive episode or dysthymia (34) 2. Non-depressed, non-psychiatric controls (14)	a) Depressed b) Manic c) Neutral	Tachistoscope presentation
Gotlib & McCann (1984)	Emotional	Students	Experiment 1 1. Dysphoric (15) 2. Non-dysphoric (15) Experiment 2 1. Induced Depression (10) 2. Induced Elation (10) 3. Induced Neutral Mood (10)	Experiment 1 & 2 a) Depressed b) Neutral c) Manic	Computer presentation

Gupta-Rogers (1999)	Emotional	Students	Experiment 2 1. Sociotropic a) Sad Mood Induction (15) b) Neutral (16) 2. Independent a) Sad Mood Induction (12) b) Neutral (10) 3. Control (neither) a) Sad Mood Induction (17) b) Neutral (16)	Experiment 2 a) Positive sociotropic b) Negative sociotropic c) Neutral sociotropic d) Positive autonomous b) Negative autonomous c) Neutral autonomous	Computer presentation Subliminal & Supraliminal
Hamilton (2003)	Emotional	Community	1. Currently MDD (23) 2. Remitted Depressed a) Negative Mood Induction (24) b) Neutral Mood Induction (23) 3. Never Depressed a) Negative Mood Induction (19) b) Neutral Mood Induction (19)	a) Negative interpersonally-oriented sociotropic b) Positive interpersonally-oriented sociotropic c) Neutral	Computer presentation
Janer (1995)	Emotional	Clinical / Community	1. MDD and/or dysthymia (20) 2. Non-depressed (16)	a) Positive b) Negative c) Nonemotional (neutral)	Computer presentation
Julian & Mhor (2006)	Classic	Clinical	1. Patients with multiple sclerosis and MDD (59)	a) Words b) Colour/word-incongruent	Not specified
Kalayam & Alexopoulos (2003)	Classic	Clinical	1. Nonpsychotic major depression (9) 2. Remitted depressed patients (13)	a) Colour/word-incongruent	Not specified

Katz et al. (1998)	Classic	Clinical	1. Attention deficit/hyperactivity disorder (89) 2. MDD or dysthymia (20)	a) Words b) Colour patches c) Colour/word-incongruent	Not specified
Kerr et al. (2005)	Emotional & Classic	Clinical	1. Bipolar Manic (14) 2. Bipolar Depressed (13) 3. Bipolar Euthymic (15) 4. MDD (17) 5. Healthy Controls (18)	a) Colour words b) Colour patches c) Colour/word-incongruent d) Negative e) Positive f) Neutral	Card presentation
Kinderman (1994)	Emotional	Clinical	1. MDD or dysthymia (16) 2. Schizophrenia or delusional disorder (16) 3. Nonpsychiatric Controls (16)	a) Negative (self-descriptive) b) Positive (self-descriptive) c) Neutral d) Nonlexical	Card presentation
Klieger & Cordner (1990)	Emotional & Classic	Students	1. Nondepressed (27) 2. Mild dysphoria (10) 3. Moderate dysphoria (10)	a) Depressed b) Neutral c) Colour/word-incongruent d) Nonlexical	Computer and projector presentation
Lemelin et al. (1996)	Emotional & Classic	Clinical	1. Major depressive episode (30) 2. Normal Controls (30)	a) Colour words b) Coloured x's c) Colour/word-incongruent	Computer presentation
Lemelin et al. (1997)	Classic	Clinical	1. Major depressive episode (moderate to severe) (33) 2. Normal Controls (30)	a) Coloured x's b) Coloured Common Words c) Colour/word-incongruent	Computer presentation

Lepage (1997)	Emotional & Classic	Students	1. Dysphoric (24) 2. Non-dysphoric (24)	a) Self-descriptive – Positive b) Self-descriptive – Negative c) Non-descriptive – Positive d) Non-descriptive – Negative e) Colour/word-incongruent	Computer presentation
Lim & Kim (2005)	Emotional	Clinical	1. MDD (30) 2. Panic Disorder (33) 3. Somatoform Disorder (25) 4. Healthy Controls (33)	a) Physical threat b) Positive c) Negative d) Categorized Neutral (household appliances)	Computer presentation Supraliminal & Subliminal
McNeil et al. (1999)	Emotional & Classic	Clinical	1. PTSD (17) 2. OCD (26) 3. MDD (18)	a) Anxiety b) Depression c) Neutral d) Colour/word-incongruent	Computer presentation
Mogg et al. (1993)	Emotional	Clinical / Community	1. GAD (19) 2. MDD (18) 3. Normal controls (18)	a) Anxiety b) Depression c) Positive d) Uncatagorized Neutral e) Categorized Neutral	Subliminal & Supraliminal
Nunn et al. (1997)	Emotional	Clinical	1. Depressive disorder, dysthymia or adjustment disorder (24) 2. Controls (24)	a) Negative sociotropic b) Positive sociotropic c) Negative autonomous d) Positive autonomous e) Depressive f) Neutral household words	Block design

Perez et al. (1999)	Emotional	Clinical / Students	1. MDD (15) 2. Dysthymia (19) 3. Dysphoric (11) 4. Negative Mood Induction (15) 5. Control (23)	a) Nonlexical b) Depression c) Elation	Card presentation
Shapiro (2002)	Emotional	Not specified	1. MDD, dysthymia, or minor depression (26) 2. Non-depressed (34)	a) Depression b) Anxiety c) Negative Sociotropy d) Negative Autonomy e) Neutral f) Euthymic g) Positive Sociotropy h) Positive Autonomy i) Total negative j) Total positive	Not specified
Siegle et al. (2004)	Classic	Not specified	1. Major Depression (23) 2. Never-Depressed Controls (28)	a) Colour/word-congruent b) Colour/word-incongruent	Computer presentation
Silberman et al. (2007)	Emotional & Classic	Clinical	1. Depressed - Parkinson's Disease (5) 2. Non-depressed – Parkinson's Disease (16)	a) Coloured dots b) Colour/word-congruent c) Colour/word-incongruent d) Negative	Card presentation
Yovel & Mineka (2004)	Emotional	Students	1. High distress (dysphoric-anxious) (123) 2. Low distress (control) (55)	a) Anxiety b) Depression c) Neutral (non-valenced)	Computer presentation