MENTAL TIME TRAVEL IN DEPRESSION: DISENTANGLING CUE VALENCE, TEMPORAL ORIENTATION AND TASK INSTRUCTIONS

By Madelaine Clair Burley

M.A., Ryerson University, Toronto, Ontario, Canada, 2010

B.A. (Hons.), University of Waterloo, Waterloo, Ontario, Canada, 2008

A Dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Psychology

Toronto, Ontario, Canada, 2017

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research. I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Mental Time Travel in Depression: Disentangling Cue Valence, Temporal Orientation and Task Instructions

Doctor of Philosophy

Madelaine Clair Burley

Psychology, 2017

Ryerson University

Abstract

Unique patterns in thinking about events from one's personal past (episodic memory) or events that could feasibly occur in one's future (episodic FT) have been linked to depression symptoms. Extensive research has examined reductions in reporting of specific memories in depression and recent interventions are designed to improve retrieval of specific memories. Reduced positive future fluency, or reduced speeded reporting of positive events plausible in one's personal future, is another characteristic of depression. While the Mental Time Travel (MTT) Model predicts a relationship between memory for past events and future thinking, there is no consensus regarding their relation in depression. This study examined the effects of Temporal Orientation (Past, Future) and Cue Valence (Negative, Neutral, Positive) within and between Fluency and Specificity Instruction Tasks (FIT and SIT) in Depressed (*n*=44) and Never Depressed (*n*=27) undergraduate student samples. Performance on Future conditions of the SIT was significantly positively correlated with performance on Past and Future FIT task conditions. The Past conditions of the SIT correlated negatively with the Future conditions of the SIT and inconsistently with the FIT. Surprisingly, neither performance on FIT nor SIT correlated significantly with depression as assessed using the Beck Depression Inventory, second edition (BDI-II). On the FIT, more events were reported in the Positive and Negative than Neutral conditions, but there were no significant differences between Past and Future conditions. On the SIT, more specific events were reported in the Neutral and Positive conditions. Lack of positive correlations within and between conditions of the FIT and SIT suggests that these tasks involve different cognitive processes. In addition, the lack of correlation between the BDI-II and both the FIT and SIT and lack of group effects suggests that more research is needed to determine moderators of reporting fluency and specificity in depression. Implications of my dissertation include adding to the considerable support of the MTT Model and its applicability to depressed samples, highlighting the need to conduct carefully controlled studies and clarify the influence of sample selection, use of prompts and scoring criteria on effects in the literature, and translate knowledge and methodology from basic research, such as models from brain imaging studies, to inform clinical practice.

Acknowledgements

I would like to acknowledge Dr. Todd Girard, my graduate advisor, for insights regarding interpretation of the literature, decisions relating to methodology, and general advice and support, and Dr. Naomi Koerner, my internal dissertation committee member for comments on my literature review and for her helpful suggestions regarding further analyses.

I would like to thank Dr. Mark Williams, Dr. Andrew Macleod, Dr. Caitlin Hitchcock, and Dr. Catherine Crane for sharing information and materials related to tasks I used in my dissertation and for consulting with me regarding scoring of these tasks.

I would like to acknowledge Diana Jin and Ryan Kopstick for their assistance with task scoring and conceptualization.

I would like to thank my committee members, Drs. Julia Spaniol, Donna Addis, and Sarah Henstra, for their helpful comments, questions, and suggestions.

Lastly, I would like to thank my labmates, Dr. Leanne Wilkins and Dr. Matt King for discussion of methodological and interpretive issues.

Dedication

I dedicate this work to my husband, who supported me throughout this process for many long years.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

The purpose of my dissertation research was to better understand Mental Time Travel (MTT) processes in people with depression. This introduction chapter contains a summary of theoretical concepts, models and findings necessary to understand the rationale for my dissertation as well as an overview of literature reviewed in depth in other sections of this paper. Major topics in this chapter include the constructive nature of memory, the MTT Model, and prominent models of vulnerability to depression. This chapter will conclude with overviews of the empirical rationale for my dissertation and my methodology.

As demonstrated by Schacter (1999, 2001) in a series of well-known experiments, memory is fallible even when confidence in memory is high. Based on findings from healthy participants, Schacter theorized that episodic memory is a constructive process, meaning that memory for past experiences is not a literal representation of past experience, but rather the result of several processes working to retrieve and combine separate thoughts, ideas and memories into a cohesive whole. As proposed by Schacter and Addis (2007), one potential function of the constructive processes underlying memory is that they enable people to simulate and pre-experience future events, which these researchers hypothesize to be critical for planning actions and problem-solving. Future thinking (FT), also called prospective thinking, simulation of future events, or episodic FT has been defined as the capacity to mentally “project” oneself forward to experience events that one imagines may happen in the future (Atance & O’Neill, 2001). Consistent with Schacter and Addis’ constructive memory hypothesis, a large body of literature suggests that similar constructive processes are shared between thinking about past experiences and future experiences (Hassabis & Maguire, 2007; Schacter et al., 2012).

Based on reviews of functional magnetic resonance imaging (fMRI) studies (Addis, Wong, & Schacter, 2007; Buckner & Carroll, 2007; Raichle et al., 2001; Szpunar, Watson, & McDermott, 2007), case studies of patients with hippocampal lesions (Hassabis, Kumaran, Vann, & Maguire, 2007; Tulving, 1985) and behavioural experiments in healthy samples (D'Argembeau & Van der Linden, 2004, 2006; Spreng & Levine, 2006), episodic memory and FT share many characteristics and appear to rely on the same neural structures. These findings have led to the conceptualization of memory and FT as MTT in opposite temporal directions, labeled the MTT Model (Suddendorf & Corballis, 1997, 2007). As depicted in Figure 1 (created by Suddendorf & Corballis, 2007), the MTT Model predicts that patterns of performance on tasks assessing the same cognitive processes across past and future will be mirrored and highly correlated.

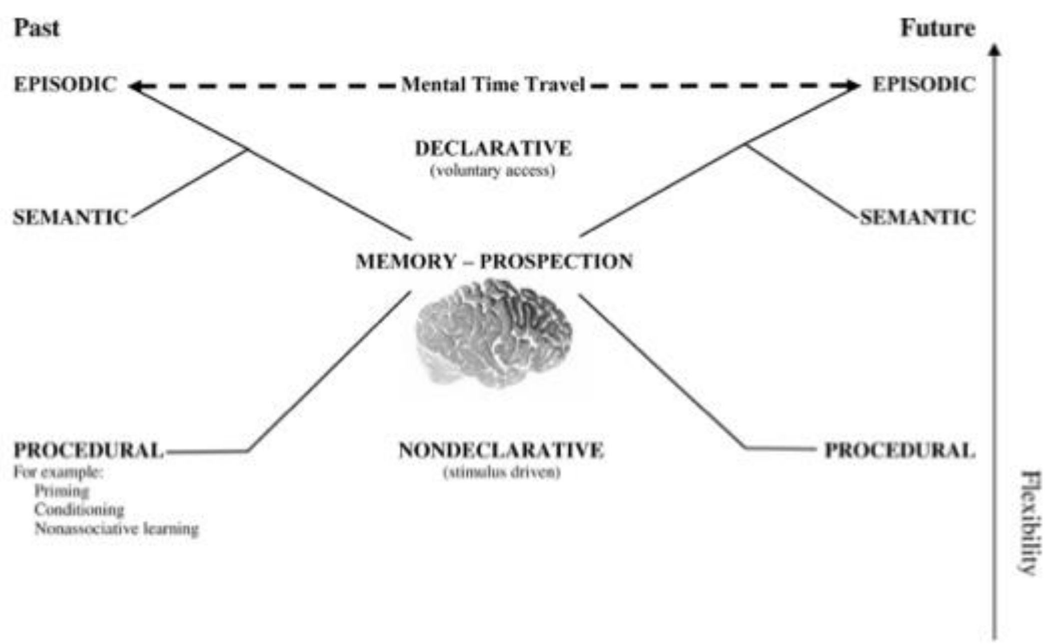


Figure 1. The Mental Time Travel Model (Suddendorf & Corballis, 2007; reprinted with permission).

Few studies have tested the MTT Model in clinical samples, although different patterns of performance on tasks assessing processes falling within the scope of the MTT Model have been reported. Understanding the processes mediating MTT is important to understand cognitive processes in everyday situations and healthy populations, but some patterns of performance on MTT tasks, such as reductions in reporting specificity and positive fluency, may also be linked to the development or maintenance of psychological disorders, including depression.

Major Depressive Disorder (MDD), commonly called depression, is a highly prevalent disorder, affecting the well-being of many. In 2004, depression was estimated to be the leading cause of disability and disease burden in the world (WHO, 2004). In 2012, depression was the most common psychological disorder in Canada: 4.7% of Canadians met diagnostic criteria for a major depressive episode during the previous 12 months and 11.3% of Canadians (approximately 3.2 million) reported retrospectively that they had symptoms consistent with a depressive episode at least once during their lifetime (Pearson, Janz, & Ali, 2013). One reason why the lifetime prevalence and burden is so high is because depression has a high rate of relapse: over a 15-year period, 85% of a large sample with remitted depression had at least one additional major depressive episode (most had several). The strongest predictor of future depression is number of previous episodes, suggesting that among people with depression, remission is associated with chronic vulnerability factors (Mueller et al., 1999; Solomon et al., 2000). Hollon, Thase and Markowitz (2002) reported that only half of all people respond to any single intervention for depression and only one third of people who receive treatment eventually meet criteria for full remission. Data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study echo the findings of Hollon and colleagues in terms of relapse and response to treatment rates, but suggest in addition that patients who attain remission have better outcomes than do patients who merely respond to treatment, reflecting that the latter still have significant symptoms following a full course of treatment (Rush et al., 2006). In summary, low rates of treatment response and high relapse rates underscore the need to improve existing treatment as well as identify new targets for treatment.

Many mediators and moderators associated with the development and maintenance of depression have been identified, including genes, neuroanatomical changes, sleep disturbances, life stressors, avoidant behaviours, dysfunctional attitudes, cognitive biases and cognitive impairments (as summarized by Garcia-Toro & Aguirre, 2007). However, high relapse rates of depression (Mueller et al., 1999) suggest that existing models of depression, upon which current empirically supported treatments are based, do not fully account for factors that moderate depression relapse and processes mediating the relation between life stressors and depression. Further research is necessary to better identify vulnerability factors in order to develop targeted treatments to reduce them with the hope of stopping the cycle of depression.

As previously alluded to, MTT processes, which include actions such as thinking about events from one's personal past (episodic memory) or events that could feasibly occur in one's future (episodic FT) have been linked to depression symptoms. Relative to samples without depression, different patterns of performance have been noted on two main types of tasks assessing declarative aspects of MTT (see Figure 1) in samples with depression. First, on tasks with instructions to report a specific memory in response to a word prompt, depressed samples report fewer specific memories relative to non-specific memories (reduced memory *specificity*). Second, on tasks where the instructions are to list as many positive or negative future events as possible within a limited time period, depressed samples report a decreased ratio of positive to negative future thoughts (reduced positive *fluency*). In addition to differing temporal focus (past or future) between tasks used to characterize MTT heuristics in depression, other aspects of task design differ in several ways, including the nature of trial cues, scoring criteria and emphases of reporting instructions within each task. Design features of the tasks used to characterize MTT in depression are the focus of my dissertation, so I have summarized and labeled the aspects of MTT task design I studied. Throughout this thesis, the term *task instructions* is used to describe global task design considerations including general focus or purpose of the task, nature of cues, number of trials, instructions to participants, and scoring procedures such as scoring criteria and indices of performance recorded during the task. As such, tasks with instructions to report a specific event in response to a word prompt are referred to as *specificity instruction tasks*, and tasks where the instructions are to list as many positive or negative events within a limited time as possible are referred to as *fluency instruction tasks*. The term *temporal orientation instructions* will refer to explicit instructions provided to participants either throughout a task or trial within a task to direct their focus to a specific temporal orientation (past or future). As a first step toward developing a more comprehensive model of the role of declarative MTT heuristics in depression that can explain distinct patterns of findings between fluency and specificity instruction MTT tasks, my dissertation study examined the influence of temporal orientation instructions (past or future) on performance across different valence conditions (positive, neutral and negative) within established specificity instruction and fluency instruction tasks.

My project improved on previous methodology by including past and future temporal orientation instruction conditions and a neutral valence condition within existing fluency and specificity instruction tasks to allow for a more nuanced consideration of the nature and scope of MTT differences in depression. Unlike previous studies, I examined performance on both specificity and fluency instruction tasks in both depressed and never depressed samples. By matching the temporal orientation aspect of the instructions and valence conditions between these tasks, I aimed to determine whether low specificity and low positive fluency in depression are linked to temporal orientation or other task parameters. A secondary aim was to explore the relations between performance variables from the fluency and specificity instruction tasks and measures of depression.

In the sections that follow, I will review research on MTT in healthy and depressed samples, outline research questions and hypotheses to be addressed with my dissertation and methodology for my study. The literature review will begin with an overview of memory specificity research and the Capture and Rumination, Functional Avoidance and eXecutive functioning (CaR-FA-X) Model (Williams et al., 2007), which purports to explain the relation between memory specificity and depression. Next, I will review findings related to future fluency in depressed samples and the MTT Model, including discussion of the cognitive processes and neural systems thought to underlie MTT. This review is important for understanding the broader impetus for my dissertation. That is, an ultimate implication supporting the rationale for my study was that understanding the processes contributing to depression-related differences in performance on MTT tasks will help to refine models of MTT in depression and their application to intervention development. The more specific goal of my study was to first address the relevance of methodological differences regarding fluency and specificity instruction tasks used in depression research. Thus, I have summarized inconsistencies between findings related to fluency and specificity in depression, the MTT Model, and predictions of the CaR-FA-X Model. I argue that methodological characteristics of specificity and fluency instruction tasks used in depression research may explain these inconsistent findings and explanations and propose methodology to begin to address this hypothesis. Lastly, I have discussed potential implications of this study for understanding performance on MTT tasks in depression, including implications for models of MTT in depression as well as interventions designed to increase memory specificity in depression.

Overgeneral Autobiographical Memory in Depression

The finding of reduced memory specificity, also called Overgeneral Autobiographical Memory (AM), has been extensively documented in depressed samples. In the following section, I have provided an overview of the existing operational definition of reduced memory specificity in depression as measured using specificity instruction tasks and summarize evidence suggesting that Overgeneral AM in depression is a distinct trait-like phenomenon reliably associated with depressive symptomatology.

As demonstrated first by Williams and colleagues (Moore, Watts, & Williams, 1988; Williams & Broadbent, 1986; Williams & Dritschel, 1988; Williams & Scott, 1988), samples of people who had previously attempted suicide generated fewer specific memories in response to cue words using past-oriented specificity instruction tasks than psychiatric and healthy control samples. Since then, it has been demonstrated that people with depression reliably report a lower proportion of specific memories relative to non-specific memories than do people without depression (Van Vreeswijk & de Wilde, 2004). Specific events are operationalized by Williams and Broadbent (1986) as events occurring at an identifiable place and time, with a total duration of less than a day. The most common examples of non-specific, or ‘overgeneral’ memories, as they have been called in the literature, are *categorical* memories, consisting of events that occur repeatedly (“walking my dog”), and *extended* memories, which are memories with a duration lasting longer than a day (“the summer of my senior year”; Williams, 1996). Based on this operational definition, Overgeneral AM can reflect a high number of reported non-specific AMs and/or a low number of specific memories. Overgeneral AM is argued to be a unique and specific impairment or heuristic bias in depression because memory specificity indices on specificity instruction tasks account for significant variance in depressive symptoms even when other cognitive processes that could explain the relation, such as lower intelligence, processing speed, semantic fluency or impairment of other forms of memory, are either absent, statistically controlled or matched across control and clinical groups (Raes et al., 2006; Park, Goodyer, & Teasdale, 2002; Wessel, Merckelback & Dekkers, 2002). Moreover, although Overgeneral AM has been reported in samples with anxiety disorders without depression, Overgeneral AM is uniquely predictive of depressive, but not anxious symptoms (Rawal & Rice, 2012).

Many studies indicate that the effect size of the difference in memory specificity between depressed and healthy samples is large. Williams et al. (2007) calculated that the mean effect size based on 11 studies comparing memory specificity between samples clinically diagnosed with MDD and matched controls was $d = 1.12$, a large effect (Cohen, 1988). As noted in the meta-analyses below (Sumner, Griffith & Mineka, 2010; van Vreeswijk & de Wilde, 2004), Overgeneral AM is most marked in samples with clinically diagnosed MDD, although modest relations have also been noted between memory specificity and dysphoria in samples who have never been diagnosed with MDD. Sumner, Griffith, and Mineka (2010) conducted a meta-analysis including findings from 15 studies published between 1993 and 2008, 9 with participants who had been diagnosed with depression and 6 with participants without clinical diagnoses of depression but whose depressive symptomatology had been assessed, allowing examination of diagnosis as a potential moderator. Sumner and colleagues reported that Overgeneral AM accounted for 1-2% of variance in depressive symptoms over time above and beyond baseline levels of depression. Clinical diagnosis significantly moderated the predictive relation between memory specificity and depression symptoms at follow-up: the predictive relation between number of specific AM and depression (without including baseline depression as a predictor) was stronger in clinical samples (standardized β =−.37, p <.0001), although the relation was trending toward significance in the non-clinical samples as well (standardized β =−.10, p =.10). Age was a trending moderator (trend toward stronger relation between Overgeneral AM and depression strengthening with age), and clinician-rated depression measures were more predictive than self-rated measures. A meta-analysis by van Vreeswijk and de Wilde (2004) yielded similar results: based on their analysis of 14 studies, van Vreeswijk and de Wilde concluded that depressive symptoms are associated with Overgeneral AM and reported that the association is stronger among clinically diagnosed samples, but noted that there was no clear consensus regarding the role of current mood state. Recent studies have shown that measures of memory specificity are stable over time in both samples with and without a history of depression, predict severity of future depressive symptoms in samples with remitted depression, and that Overgeneral AM can predict initial onset of depression under certain conditions (Rawal & Rice, 2012; Stange, Hamlat, Hamilton, Abramson, & Alloy, 2013; Sumner et al., 2014). In sum, these findings indicate that Overgeneral AM is a trait-like phenomenon in depression. The strength of evidence of Overgeneral AM and its stability have prompted numerous studies seeking to explain mechanisms underlying the phenomenon as well as to examine its potential as a mediator or moderator of depression. These efforts culminated in the development of a highly influential model, the CaR-FA-X Model (Williams et al., 2007), the claims of which are currently under investigation and will be discussed in the following section.

CaR-FA-X Model

Following an in-depth synthesis of extensive literature, Williams and colleagues (2007) proposed a model to explain how differences in processes mediating retrieval of memory for personal experiences contribute to the high rate of relapse of depression. Throughout this section, I will outline the theoretical framework of the CaR-FA-X Model proposed by Williams et al., discuss its influence on contemporary depression research and highlight inconsistencies between predictions from the model and patterns of performance associated with different valence conditions of specificity instruction tasks.

In 2007, Williams and colleagues proposed a model to account for the finding of Overgeneral AM in depression by linking the findings to a model of retrieval of self-relevant memories proposed by Conway and Pleydell-Pearce (2000). As depicted in Figure 2, the Self-Memory System (SMS) Model developed by Conway and Pleydell-Pearce (2000) purports that memories are stored using a top-down hierarchical structure. Therefore, when volitionally retrieving declarative memories of specific autobiographical events, a hierarchical search is used, wherein a general theme (e.g., ‘work’) is chosen, followed by a life era for the theme (‘past five years at work’), then categories of events within that theme and era (‘parties at work in past five years’), and finally specific memories are retrieved (‘office holiday party two years ago’). Applying this framework to explain Overgeneral AM, Williams and colleagues (2007) proposed three processes that interfere with memory retrieval in depression: Capture and Rumination, Functional Avoidance and impaired eXecutive control (The CaR-FA-X Model; see Figure 3).

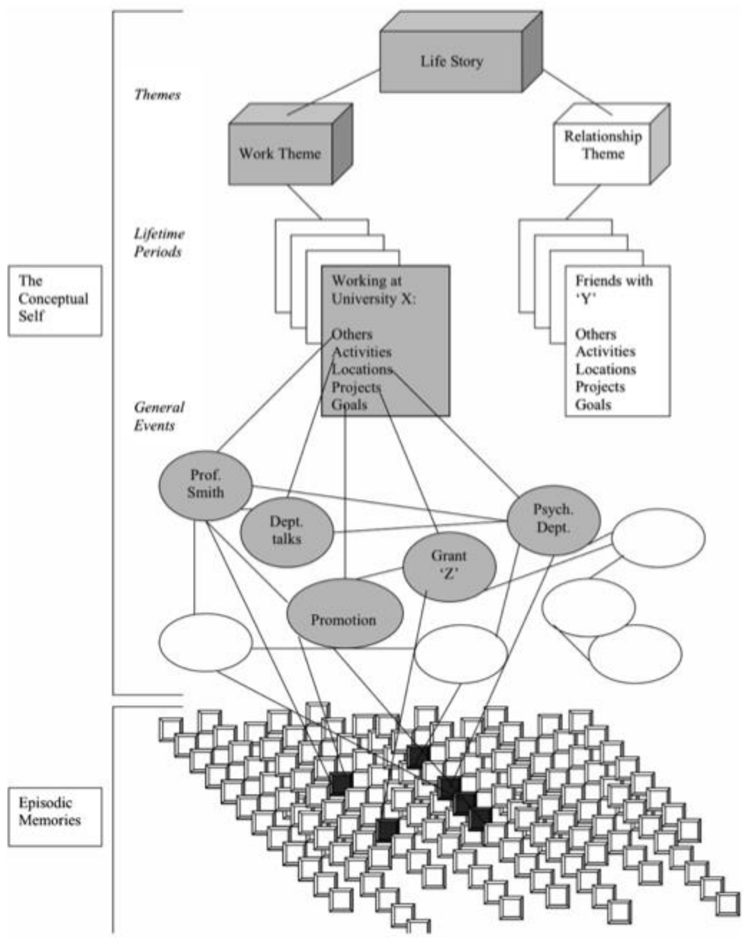


Figure 2. The Self-Memory System Model of intentional memory retrieval (Conway, 2005; reprinted with permission).

When cued to retrieve a specific memory, Conway (2005) proposes that as the hierarchical search process progresses toward retrieval of a specific event, called generative retrieval by Conway, sensory and perceptual memories are retrieved. Williams et al. contend that the sensory and perceptual aspects of specific memory retrieval elicit emotional responses and cognitions consistent with re-experiencing the event. Therefore, when people with depression are cued to retrieve specific memories, re-experiencing the event causes negative cognitions and emotions become activated (Capture) as well as the maladaptive cognitive response called Rumination (Wisco & Nolen-Hoeksema, 2009). Williams et al. theorize that rumination is so emotionally salient that it dominates attentional resources and that people with depression are not able to employ eXecutive control to refocus on memory retrieval. Consistent with recent literature, Williams and colleagues propose that over time, people with Overgeneral AM adopt a strategy of curtailing the search process for specific memories at the level of general events (see Figure 2) to avoid cueing highly aversive negative thoughts. This strategy is hypothesized to relieve distress in the short-term (Functional Avoidance); however, memory specificity has been demonstrated to contribute to successful problem-solving. Problem-solving (among several other processes) is thought to mediate the relation between life stressors and onset of depression; therefore, avoidance resulting in reduced memory specificity is hypothesized to result in ongoing impairment in problem-solving, increasing the risk of depression onset, maintenance and remission. Several of these premises have been supported by empirical studies using analogue samples and quasi-experimental studies using depressed samples and research is ongoing (reviewed by Sumner, 2012). Based on a review of 38 studies testing aspects of the CaR-FA-X Model, Sumner (2012) reported that most aspects of the model have robust empirical support.

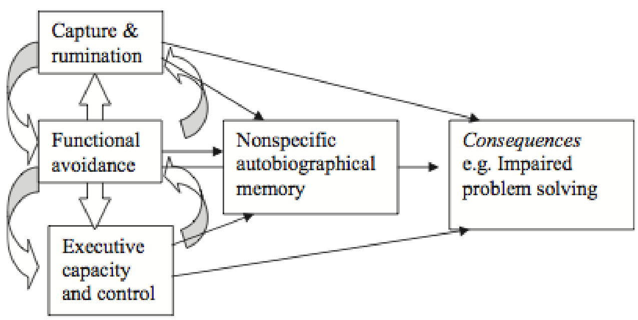


*Figure 3.*The Capture and Rumination, Functional Avoidance and eXecutive functioning (CaR-FA-X) Model of depression vulnerability (Williams et al., 2007; reprinted with permission).

Recently, a *MEmory Specificity Training program* (MEST) has been developed by researchers core to the development of the CaR-FA-X Model and collaborators (Neshat-Doost et al., 2012; Raes, Williams, & Hermans, 2009). Two clinical trials, one using a within-subjects design and the other a waitlist-control sample indicate that the protocol increases memory specificity in depressed samples and that these increases are correlated with reduced symptoms of depression. On the basis of these promising early findings, the first randomized controlled trial for this protocol was registered with the National Institutes of Health in 2013 (http://clinicaltrials.gov/show/NCT01882452). Nonetheless, several findings inconsistent with the Functional Avoidance hypothesis of the CaR-FA-X model have yet to be explained. That is, the Functional Avoidance hypothesis predicts that negative memories would be avoided more than neutral or positive memories, resulting in lower memory specificity when prompted to recall negative events (see discussion by Williams et al., 2007, p. 135). Individual studies examining whether valence moderates memory specificity have been inconsistent. However factor analytic studies, correlational findings and meta-analyses suggest that reduced specificity is found across positive and negative conditions of specificity instruction tasks with varied parameters in depressed samples (Griffith et al., 2009; Sumner, 2012; Sumner et al., 2010; van Vreeswijk et al., 2004; Williams et al., 2007).

Williams et al. (2007) proposed that the lack of valence differences may reflect the influence of a combination of negative self-schemas and direct, rather than hierarchical memory retrieval processes occurring in depression. Direct retrieval, suggests Conway (2008), is not volitional and occurs when memories are spontaneously cued by internal or external stimuli, resulting in retrieval of memories without hierarchical search. Williams et al. contend that because depression is associated with broad negative self-schemas (Beck et al., 1979; Segal, 1988) categorical cues that are designed to facilitate retrieval of positive memories (e.g. ‘summer break’) or positive memories themselves (“that day on the beach with my family”) may also cue direct retrieval of negative memories or act as cues for negative memories (“critical comments from my family") and therefore retrieval of positive specific memories as well as negative specific memories is avoided.

If, as proposed, the unexpected reduced specificity of positive AMs is attributable to functional avoidance via curtailed hierarchical search processes and/or direct retrieval, it is reasonable to hypothesize that a similar pattern of diffuse avoidance will also occur during other tasks with similar potential to result in negative cognition during declarative MTT. Unlike specificity instruction tasks, reduced reporting of positive events but comparable reporting of negative events on fluency instruction tasks by depressed samples is inconsistent with a functional avoidance hypothesis. In the following section, I review research on performance on fluency instruction tasks in depression, MTT in healthy samples, and studies comparing performance on specificity and fluency instruction tasks to support the hypothesis that there may be a mechanism besides functional avoidance that mediates the relation between depression and fluency task performance. Understanding the task factors accounting for differences between specificity and fluency instruction tasks is an important first step toward determining whether the CaR-FA-X Model can also account for performance on fluency instruction tasks in depression or whether this pattern suggests other mechanisms of depression vulnerability.

Impaired Positive Future Thinking in Depression

Interest in FT in depression began when it was noted that people who had recently attempted to commit suicide reported fewer positive events ("I will have brunch with Katie in November") when instructed to report as many FTs as possible in one minute than people without depression (i.e. a fluency instruction task). Conversely, they generated a similar number of negative FTs given the same time limit, suggesting specific impairment for generation of positive events (MacLeod, Rose, & Williams, 1993; for a recent review, see Szpunar, 2010). In the future-oriented fluency instruction task used in these studies, participants were prompted to list as many important or trivial events as they could in one minute for each of three different time periods in the future (1 week, 1 year, 5-10 years); in the positive condition, they were instructed to report things they knew or thought might happen to which they were looking forward, and in the negative condition, they reported events that they knew or thought might happen that they were worried about or to which they were not looking forward (as described by MacLeod, Tata, Kentish, & Jacobsen, 1997). Subsequent research extended these findings to samples with depression without suicidality as well as those with mild depression (MacLeod & Salaminiou, 2001). Interestingly, a sample of individuals with panic disorder demonstrated the reverse pattern of results, generating more negative FTs than controls, but similar numbers of positive FTs, suggesting that impairment in positive FT is specific to depression (MacLeod et al., 1997). Like Overgeneral AM, impaired positive FT is not attributable to differences in IQ or verbal fluency between depressed and healthy samples, or differences in anticipated positive emotions related to FTs in depression (MacLeod & Salaminiou, 2001).

MacLeod and colleagues also developed tasks designed to assess MTT in depression. MacLeod, Tata, Kentish and Jacobsen (1996) examined past and future fluency in healthy, depressed, and anxious samples and found reduced fluency across past and future in depressed, but not anxious or healthy participants. Importantly, this pattern was found across both past and future, and positive and negative valence conditions. Although this study provides support for the applicability of the MTT Model for fluency tasks in that patterns were mirrored across temporal directions, MacLeod and colleagues did not replicate the finding of reduced positive fluency and no possible explanations for this were provided by the authors. Replication of this study with the addition of a neutral valence condition could help by allowing the influence of emotion intensity rather than just emotion valence to be examined.

In contrast to Williams and colleagues' CaR-FA-X Model, which proposes that biases in memory serve the function of avoiding negative emotions, the finding of impaired positive FT in depressed samples is inconsistent because it is difficult to reconcile impairments in the positive condition of a task with normal performance on the negative condition of a task with avoidance. If avoidance strategies are adopted by people with depression on one task, why are they not adopted on others that also present opportunity to elicit negative emotion? Based on this inconsistency, it is probable that the cognitive mechanisms underlying typical assessment of reduced positive FT and Overgeneral AM in depression are different in one or more ways. However these tasks have rarely been studied concurrently using within-subjects designs among samples with depression (an exception is Sarkohi et al., 2011; discussed in this document on p. 30) and the mechanisms underlying differences between them as well as their relation to depression have not been articulated in the CaR-FA-X or other models of depression. As will be illustrated in the coming section, research on MTT in healthy adults may provide insight into possible neural and cognitive mechanisms underlying this inconsistency, but these basic research insights on the processes and factors underlying performance on different types of MTT tasks are not reflected in task design and theory about MTT in depression.

Mental Time Travel in Healthy Samples

In depression research, AM and FT have most often been studied separately (for exceptions see Addis, Hach, & Tippett, 2016; Anderson & Evans, 2015; Dickson & Bates, 2006). However behavioural studies using healthy samples suggest that thinking about the past (AM) and thinking about the future (FT) are highly interdependent and phenomenologically similar processes (see reviews by Atance & O’Neill, 2001; Szpunar, 2010). In healthy samples, findings related to AM are typically replicated in FT and vice versa; these findings have led to the characterization of both AM and FT as MTT processes, dubbed the MTT Model (see Figure 1; Suddendorf & Corballis, 1997). I review important findings related to future and past thinking using healthy samples that may be relevant to MTT in depression related to temporal orientation, valence and instruction emphases, which I examined for my dissertation. My aim is to demonstrate the rationale for my study by highlighting the potential relevance of basic MTT research to MTT research in depression, emphasizing the importance of understanding MTT task factors and the cognitive processes supporting them and illustrating what may be gained by translating findings and methodology from studies using healthy participants such as the MTT Model to studies of MTT in depression. In particular, I note factors that are not routinely assessed or manipulated in fluency and specificity instruction tasks that could explain some of the inconsistencies in findings between these tasks. The importance of this is that to understand the patterns of performance in MTT in depression, assessment of past and future MTT is necessary. Following an overview of MTT research in healthy samples, I will re-state the importance of these findings and support the rationale for my dissertation.

Using modified versions of specificity and fluency instruction tasks used in depression studies, D’Argembeau and Van der Linden (2004) examined the relation between phenomenological characteristics of FT and episodic memory in a sample of healthy adults. Participants were instructed to retrieve specific events from their past or generate/construct specific FTs that they expect could reasonably happen to them; participants also rated phenomenological aspects such as perceptual detail, valence and emotional intensity, clarity of spatial information, visual perspective and estimated temporal distance of events from the present. D'Argembeau and Van der Linden reported that memories were described as richer and more vivid on average than FTs. However, for both past and future events, positive events were associated with more vivid sensory and emotional phenomenological experiences, often termed re-living or pre-experiencing, and events that were closer to the present included more detail, and were more commonly associated with first-person perspective. D'Argembeau and Van der Linden (2006) subsequently reported that individual differences in visual imagery ability were associated with more vivid and rich sensory re- or pre-experiencing of events, and that trait emotional suppression was associated with less vivid MTT experiences. The finding of increased vividness of positive events might help explain valence effects in AM and FT in depression if vividness experiences differ between tasks used to assess AM and FT, but the potential role of event vividness has never been considered. In addition, temporal distance from the present has not been explored as a contributor to MTT heuristic biases in depression, and the generalizability of the finding of increased vividness of past compared to future events and positive compared to negative events in depression is unclear. The lack of consistency of parameters between tasks used to assess AM and FT in depression and unclear generalizability of the findings of increased vividness of past positive events from healthy to depressed samples means that the precise nature of the differences in MTT in depression is unknown. Importantly, these studies show interactions between temporal distance, temporal orientation and valence in healthy samples, suggesting that trends related to vividness may be associated with healthy functioning. By comparing typical MTT specificity and fluency instruction task performance in healthy and depressed samples, a goal of this dissertation was to determine in what ways the patterns of performance are similar and different; based on the overall patterns, this design will allow me to rule out potential confounds in the literature and suggest, based on the literature on healthy adults, which mechanisms may underlie MTT differences in depression. Overall, these studies suggest that there are still many unexamined potential moderators and mediators of performance on specificity and fluency instruction tasks and therefore that much more research is needed to pinpoint which moderators and mediators underlie the patterns found in depression.

In addition to behavioural studies, neuroimaging studies also suggest similarity between AM and FT in healthy populations. Schacter, Addis and colleagues (Addis, Pan, Vu, Lauser, & Schacter, 2007; Addis & Schacter, 2008; Addis, Wong, & Schacter, 2008) have conducted many fMRI studies to determine common and distinct neural pathways associated with MTT. Addis et al. (2007) used 96 nouns high in frequency, imageability and concreteness as cues and asked healthy participants to *construct* events from either the past or future related to the noun cue (at distances of one week, one year and 5-10 years), indicate when they had an event in mind by pushing a button and spend the remainder of each 20-second trial mentally *elaborating* upon that event. After scanning, cues/events generated were rated for personal significance; emotionality and perspective ratings (i.e., 1^{st} vs. 3^{rd} person) were also collected. Addis et al. (2007) reported that brain areas that were commonly activated by past and future trials during the elaboration portion of the task were consistent with areas reported to be active during AM retrieval in prior studies, whereas future event construction activated the right hippocampus uniquely; this study highlights the distinction between construction (generation) and elaboration in MTT tasks. Addis et al. (2008a) used the same data to conduct analyses to demonstrate that temporal distance and amount of detail provided affect the amount of brain activation during the task; whereas recent events correlated with activation in the right parahippocampal gyrus, distant events correlated with activation of the bilateral hippocampus, and whereas activation of the left posterior hippocampus was modulated by detail during trials tapping past and future, the left anterior hippocampus was associated only with future event processing. Addis et al. (2008) used samples of events generated by past participants in a subsequent study and asked participants to imagine future and past events using examples supplied to determine the extent to which the process of imagining accounted for differences noted between past and future conditions. Addis et al. (2008) reported that across both past and future imagining, the same regions were generally activated, suggesting that although thinking about future and past naturally elicit different degrees of imagination, the same networks are used in either direction, and thus, the same neurocognitive elements underlie both past and future thinking. Szpunar et al. (2007) replicated the results of Addis et al. (2008) and reported that emotional intensity and valence did not change the overall pattern of findings. Based on these studies, Addis and colleagues argue that AM and FT share similar underlying neural systems.

Weiler, Suchan, and Daum (2009) also examined neural networks associated with AM and FT in healthy participants using fMRI. Participants were asked to recall or construct events from either the most recent holiday season or the upcoming holiday season and verbally indicate when they had an event in mind, while in the fMRI machine. These instructions were designed to ensure that thought content was similar between past and future conditions of the task. Weiler and colleagues found that phenomenological aspects of FTs and AMs were similar, but that different cortical networks were accessed for each time orientation, and that the time course of activation was different between FT and AM. Central findings included higher activation in the right posterior hippocampus during initial AM construction (generation) in the early phase of the task, but greater activation for FT during the latter half of the task, the elaborative component. Observations by Weiler et al. (2009) converge with those reviewed above by Addis, Schacter and colleagues (Addis et al., 2007, 2008; Addis & Schacter, 2008), that construction (recalling or generating an event) is dissociable from elaboration and that construction and elaboration are differentially related to AM and FT task demands but suggest that further research is needed to understand similarities and differences between time course of activation and the effect of temporal orientation on MTT.

In summary, behavioural studies illustrate that AM and FT differ in vividness of re-/pre-experiencing and that factors such as temporal distance and valence also affect vividness of MTT (D'Argembeau & Van der Linden, 2004, 2006). Brain imaging (fMRI) research indicates that MTT is not a unitary cognitive process. The act of recalling a memory can be broken down into two distinct processes: Construction and Elaboration. Recalling the SMS Model of memory retrieval (depicted in Figure 2) and given the nature of tasks used to dissociate these processes, Construction likely involves processes supporting hierarchical search based on time period or category resulting in retrieval of a specific memory whereas Elaboration likely involves generation of episodic details and re/pre- experiencing of events (Addis et al., 2007, 2008; Addis & Schacter, 2008). Although the specific cognitive components involved in these processes are still under investigation, components that have been hypothesized to be common between re- and pre-experiencing an event could include theory of mind/self-awareness (Buckner & Carroll, 2007; Hassabis & Maguire, 2007), visual imagery and hierarchical search processes. As the tasks used to assess factors implicated in MTT are better understood, it is hoped that cognitive substrates central to both tasks will be more clearly articulated. In healthy samples, the implications of these findings include the possibility of consolidating research efforts from AM and FT into a single, clearer model of MTT, which in turn gives a more complete account of processes influential to or reliant upon MTT.

While the implications of the MTT Model for healthy samples has received considerable attention, the implications for clinical samples are less clear because few studies have sought to test whether findings supporting the MTT Model generalize from healthy samples to clinical populations. Specifically, whereas some factors linked to normal MTT heuristics, such as valence, have been examined in depressed samples, others, such as temporal orientation, variations of which elicit different levels of detail reporting, vividness and re-experiencing, have not been examined. Because re-experiencing is hypothesized to underlie functional avoidance and therefore valence effects in specificity and fluency instruction tasks according to the CaR-FA-X Model, it is important to determine whether the valence effects are strongest in temporal orientation conditions shown to increase re/pre-experiencing in previous studies. The current tasks are not able to address this question because they do not include both temporal orientations within each task, allowing for direct comparison within a single sample; therefore, by including new temporal orientation conditions my dissertation addresses an important gap in the literature.

In addition to informing MTT theory and the CaR-FA-X Model, understanding the extent to which MTT findings generalize to depressed samples has practical clinical implications. For instance, rather than disparate research teams and projects contributing to separate literatures on AM and FT, it may prove more profitable to focus investigation under the more unified framework of MTT. In addition, understanding the extent to which past and future thinking share similar processes may lead to clearer identification of disease mechanisms, better methods of assessment and new targets for treatment. In this regard, it was also important to assess the potential areas of convergence between MTT theory and current models of depression like CaR-FA-X.

Inconsistencies between CaR-FA-X Model and MTT Model

Williams et al.'s (2007) functional avoidance hypothesis from the CaR-FA-X Model proposes that depression is linked to a tendency to avoid recalling negative events by limiting search strategies in tasks requiring volitional cognitive retrieval strategies. Given evidence supporting the MTT Model that FT and AM share common cognitive processes, one might expect that similar processes and biases occurring in memory in depression exist in FT in depression. Therefore, based on the CaR-FA-X Model, it would be predicted that depressed samples would report a lower number of negative FT or similar positive and negative FT (a replication of the findings from the AM literature, though not yet fully understood) on fluency instruction tasks. As reviewed above, the robust finding of impaired positive FT on fluency instruction tasks is difficult to reconcile with these model predictions. In addition to findings inconsistent with the FA hypothesis from the CaR-FA-X Model, generalizability of predictions from the model regarding the relation between FT and Rumination (see Figure 3) in depression is also not clear. Attempts to extend findings linking Overgeneral AM and Rumination to reduced positive FT and Rumination have reported similarly confusing results. For instance, following Rumination, a depressed sample reported higher numbers of both positive and negative FTs on a fluency instruction task, suggesting that unlike Overgeneral AM, impaired positive FT may not be linked to Rumination (Lavender & Watkins, 2004). The authors of this study hypothesize that this may be due to a general priming of self-relevant schemas induced by rumination resulting in quicker identification of FTs, but this explanation is also inconsistent with a Functional Avoidance explanation. Thus, while the CaR-FA-X Model may adequately explain findings related to AM in depression, attempts to demonstrate similar relations between FT and depression suggest that important differences exist between AM and FT, or the tasks used to assess these constructs.

Similarly, although the MTT Model highlights overlap between AM and FT processes, research using depressed samples in which patterns of performance on valence conditions of AM and FT tasks differ highlights the need to better understand the tasks used to examine past and future thinking. In turn, this research will help to determine the extent of overlap between these tasks and factors that differ either in intensity (e.g. vividness) or in relative importance (e.g. construction versus elaboration) in the assessment of AM and FT in depressed samples. The apparent discrepancies between AM and FT research findings in depression suggest that the models and/or methods used may need revision.

In spite of the abundance of knowledge obtained from basic biological and cognitive studies regarding processes involved in AM and FT in healthy samples, factors linked to task performance identified in these literatures, such as vividness, are not reflected in the selection of methods and interpretation of results for studies of Overgeneral AM and FT impairments in depression. In the next section, I will discuss evidence that consistent with research on the MTT and SMS Models, tasks commonly used to assess fluency and specificity in depression likely rely on different cognitive processes and neglect to control for factors that are affecting results. I believe these factors are involved in performance on AM tasks, may be differentially affected by depression, and therefore may be relevant to the CaR-FA-X Model. Understanding factors that affect performance on these tasks will not only inform cognitive models of depression, such as the CaR-FA-X model, but also the MTT model and could one day inform treatment approaches. In the section to follow, I have summarized theoretical and methodological issues with tasks commonly used to assess MTT in depression and research on the processes involved in MTT tasks used in depression research.

Assessment of MTT in Depression

Although reduced memory specificity in depression, or Overgeneral AM has been demonstrated using several versions of specificity instruction tasks with varied parameters, the effect has been studied most often in samples with depression using the *Autobiographical Memory Test* (AMT; Griffith et al., 2012; Sumner, Griffith, & Mineka, 2010; van Vreeswijk & de Wilde, 2004; Williams & Broadbent, 1986). In the AMT participants are instructed to recall and report specific memories cued by positive, negative and occasionally neutral words. Participants are instructed and trained to report specific memories of events that can have occurred recently or long ago and may be important or trivial, and are usually given limited time (ranging from 30 seconds to 2 minutes) to retrieve a memory for each cue word. Despite being instructed to retrieve specific memories ("I had brunch with Katie on Sunday"), overgeneral memories ("Having brunch with friends") are more often reported by depressed samples. Responses are coded afterwards as either specific memories, extended memories ("the summer of my senior year"), categorical memories ("walking my dog"), semantic associations ("Christmas") or omissions. Often, latency (time until response is generated) is recorded and in some iterations of the task, participants may be given a second opportunity to report a specific memory for responses that were not specific upon first presentation of the cue.

Griffith and colleagues (2012) reviewed methodology used and psychometric issues in Overgeneral AM research with clinical samples. Griffith and colleagues suggest that task parameters, such as instructions, affect study outcome and that the effect of task parameters must be understood to properly interpret research findings. For instance, a variation of the AMT with minimal instructions can reduce ceiling effects in healthy or analogue samples because without the instruction to be specific, their natural tendency may be to retrieve fewer specific or more overgeneral memories. In a study by Debeer, Hermans, and Raes (2009), significant relations between depression, rumination and Overgeneral AM were found for scores from a minimal instruction condition, but not from a condition using the typical memory specificity task instructions in a non-clinical sample. The results of the study by Debeer and colleagues (2009) demonstrate that the typical instructions of the AMT may inadvertently be eliciting more specific responses from participants. Importantly, this could indicate that the AMT instructions reduce task sensitivity, or that depressed samples are less adept at modulating their responses based on task instructions. Direct comparison of performance on the typical and minimal instruction AMT in a depressed sample could test these hypotheses. Griffith et al. reviewed studies using alternative measures of Overgeneral AM such as the Test Episodique de Mémoire du Passé autobiographique (TEMPau; Piolino, Desgranges, & Eustache, 2009) and the Sentence Completion for Event of the Past Test (SCEPT; Raes et al., 2007); although studies using these measures generally supported patterns found using the AMT, minor differences in task parameters moderated outcomes. Griffith et al. concluded from these studies that specific parameters of AM tasks significantly affect study results and may contribute to inconsistent results between studies.

Griffith et al. (2009, 2012) examined the reliability and validity of the most commonly used test, the AMT. Across three healthy samples, the AMT was best modeled by a one-factor solution, suggesting that responses are related to a unitary construct, regardless of valence (Griffith et al., 2009). The internal consistency reliability of the traditional AMT was above a standard acceptability threshold for research using one method of assessing reliability (reliability point estimate, calculated using the method outlined by Raykov, Dimitrov and Asparouhov (2010) was .79, 95% CI: .74-.84), but not based on Chronbach's Alpha (.54), and the minimal instruction AMT had poorer reliability (with Raykov et al.'s method: .64; 95% CI: .57-.72; Chronbach's Alpha: .53). The test re-test reliability of alternate forms of the original AMT indicated correlation of medium magnitude over time. Griffith et al. (2012) noted that the test re-test reliability variance is consistent with literature demonstrating the moderating effect of current mood state on memory specificity or could be due to task instructions directing participants to respond in the moment (i.e., participants may recall different memories in response to the same cue on different days).

Although the AMT task is well-studied and commonly used, the extent to which sub-processes of AM may influence performance has never been considered, nor has the extent to which such processes have been confounded in MTT research in depression. The focus of my dissertation was to understand the effect of specificity and fluency instructions and temporal orientation instructions on task performance, but the implications of this study could include suggesting impairment on one or more specific sub-processes involved in retrieval/generation of events. To illustrate, I have briefly discussed predictions about sub-processes involved in specificity and fluency instruction tasks based on neuroimaging findings. For instance, comparison of neural activation during the first ten seconds of episodic retrieval to the latter twenty seconds by Addis and colleagues (2007) suggests that two distinct processes take place during the retrieval of autobiographical events, which they dubbed construction and elaboration. The AMT task demands are strikingly similar to the task used by Addis and colleagues, but AMT task scoring does not distinguish between construction and elaboration processes occurring throughout the task: participants are instructed to both *construct* a memory in response to a cue word and also *elaborate* with specific details for times ranging between 30 seconds and two minutes. Based on the similarities between the instructions of the AMT task and the task used by Addis et al., both processes likely contribute to task performance, but the unique contributions of each process to performance on AMT have never been examined. Given the timeline for transition between construction and elaboration put forth by Addis et al., and the length of each AMT trial, it is reasonable to hypothesize that the majority of cognitive activity occurring during the task is likely to be elaboration, but this question has not been examined empirically.

In contrast, task parameters of the *Personal Future Task* (PFT; Macleod, Rose, & Williams, 1993), which is the most commonly used fluency instruction task in depression, appear weighted to favour construction processes. The PFT assesses the ease with which participants can report positive (‘things you are looking forward to’) and negative (‘things you are not looking forward to’) events across future time periods (typically 2-5) ranging from the next 24 hours to the next 10 years; participants are given 30 to 60 seconds to provide as many examples for each temporal period and valence condition as they are able. The PFT was based on a common neuropsychological measure, the verbal fluency task (FAS); the FAS is thought to assess executive functioning. During the FAS participants are instructed to list as many words that start with a given letter as they can for 30 to 60 seconds with greater numbers of words reported being associated with better executive functioning ability. The FAS has previously been used as a control task in studies using the PFT to demonstrate that impairments noted in clinical samples on the PFT are not fully accounted for by underlying differences in verbal fluency. Consistent with the FAS, instructions on the PFT emphasize generation of events for the duration of the task; because participants are instructed to report as many events as possible during the PFT, construction processes, which are associated with generation of events may be active throughout each trial of the task. In the PFT, participants are not explicitly directed to elaborate on the events they generate; indeed, elaboration would be counterproductive since the task is time limited. Thus, different task instructions emphasize different underlying cognitive and neural processes, which may account for some of the apparent discrepancies between findings regarding AM based on specificity (i.e., AMT) and FT based on fluency (i.e., PFT) instruction tasks in depression.

Indirect evidence from at least two studies (D'Argembeau, Ortoleva, Jumentier, & Van der Linden, 2010; Sarkohi, Bjärehed, & Andersson, 2011) supports the hypothesis that discrepancies in MTT research in depression can be accounted for by uncontrolled task parameters differing between specificity and fluency instruction tasks, further adding to the evidence of the need to reconsider the design of fluency and specificity instruction tasks and refine existing models of MTT in depression. D'Argembeau and colleagues (2010) demonstrated that MTT fluency, specificity and elaboration tasks are supported by separate neuropsychological processes and that understanding the processes underlying performance is important to understand heuristics in MTT and also similarities between FT and AM. These authors examined relations between MTT task performance and neuropsychological processes including working memory, executive control, relational processes, visual-spatial processing, self-consciousness and subjective sense of time. D'Argembeau and colleagues used modified specificity and fluency instruction tasks as well as an Episodic Details task adapted from Hassabis et al. (2007) to characterize different aspects of MTT in healthy adults. Each of the three MTT tasks used included both past and future conditions with matched instructions, thereby equating temporal direction across all tasks (six conditions total). During the Episodic details task participants were given sentence prompts ("recall/imagine the last/next time you met/will meet a friend"; "recall/imagine something you did/will do during your last/ next vacation") and instructed to vividly imagine themselves in that scenario and report as many episodic details as possible to provide an indication of the number of episodic details participants are able to generate independent from their ability to generate/recall their own scenarios. Main findings highlighted by D'Argembeau et al. included reliance on executive processes across several MTT tasks for both past and future, greater reliance on visual-spatial processing for sensory details for future events only, self-consciousness and future oriented thinking were correlated with number of sensory descriptions for future events, self-consciousness was associated with subjective pre-experiencing for future events, and executive processes, visual-spatial processing and future-orientation independently predicted number of sensory descriptions reported for future events. Consistent with prior studies conducted by this group, past task conditions were associated with a greater number of responses and subjective measures of re-/pre-experiencing than future conditions. Overall, this study highlights the need to carefully consider task design in studies of MTT, as it suggests that although AM and FT share variance, cognitive processes are involved to different extents in each. An important limitation of this study is that across MTT tasks, valence was not assessed, although typical use of MTT tasks in clinical samples includes positive and negative valence conditions and valence has been noted as a potential moderator of performance. Finally, whereas previous studies have focused more on highlighting the similar patterns of results across temporal orientation conditions within and between specificity and fluency instruction tasks in healthy samples, this study provides valuable data concerning some different relations between these tasks with neuropsychological measures. By extension, the different nature of findings reported between these tasks in depression suggests that factors underlying MTT task performance are differentially affected in depressed populations. Studying factors associated with specificity and fluency instruction tasks in depression could therefore be informative to the MTT Model by suggesting a new mood-related moderator or mediator of specificity and fluency tasks/processes.

Sarkohi et al. (2011) compared the AMT and PFT using a within-subjects design in a sample of 88 individuals with mild to moderate depression. The sample met criteria for depression diagnosis on the basis of a structured interview and depression severity was characterized using the Montgomery Åsberg depression rating scale (MADRS-S; Svanborg & Åsberg, 2001) and Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). In the PFT, participants generated plausible future events, but also rated their belief in the likelihood of these events and the emotional valence associated with their thoughts. The AMT version used in this study consisted of 36 emotional words, balanced across positive, negative and neutral valence and administered in two sets; participants had one minute to report a memory for each word. In this sample, positive FTs were only moderately correlated with number of specific positive AMs retrieved (r = .23) and positive AMs were positively correlated with negative AMs (r =.39). Both positive and negative specific AMs were negatively correlated with indices of Overgeneral AMs (rs > -.27). In contrast to what has been found in other studies, depression and anxiety self-report measures were uncorrelated with PFT or AMT scores. Given that all participants in the study currently had depression and there were no control groups, restricted range may have limited the ability to detect a correlation. As discussed in previous sections, the MTT model proposes that the same cognitive processes occur during AM and FT. However, in Sarkohi et al.'s (2011) study, negative AM scores were not correlated with positive or negative FT scores. Sarkohi and colleagues suggest that one possible explanation behind this finding is that the PFT and AMT assess different processes. Further, the specificity and fluency instruction tasks used by Sarkohi et al. are confounded with temporal orientation. That is, although the lack of correlation between AMT and PFT performance might be due to differential underlying cognitive processes associated with specificity versus fluency instructions, their design cannot rule out the possibility that the differences might be attributed to differing temporal orientation instructions between the tasks. In addition, the absence of a healthy control group makes it difficult to determine whether the relations described, and thereby underlying mechanisms, are uniquely associated with depression or are consistent with the general population. This study highlights the need to systematically compare the specificity and fluency instruction tasks in depressed samples to determine what factors (e.g., temporal orientation, valence) account for different patterns between the tasks, and the extent to which each is related to mood or depression. In sum, although the MTT Model highlights commonalities between AM and FT, this study suggests that separate mechanisms may underlie findings related to FT and AM tasks in depression. The extent to which AM and FT are related is an important question because if they are linked, they may be targeted using the same treatment approaches, whereas if they are not linked, they represent distinct vulnerabilities that warrant separate treatment considerations.

A final criticism concerning assessment of MTT in depression pertains to a task design issue affecting both specificity and fluency instruction tasks: lack of consistent inclusion of neutral emotion conditions within the tasks. Memory research using healthy samples indicates that positively and negatively valenced events are recalled more easily than neutral events (Reisberg & Hertel, 2004) and that negative images are recalled significantly more than both neutral and positive ones (D’Argembeau & Van der Linden, 2005). As noted above, Sarkohi and colleagues (2011) included a neutral cue condition in their specificity instruction task, but no neutral condition was included on the fluency instruction task limiting comparison and relations between performance on the neutral condition and other task variables and measures was not reported. Some previous studies of memory specificity that have included a neutral emotional condition suggest that contrary to other memory tasks, specificity instruction tasks demonstrate a pattern of greater specificity in response to neutral relative to emotional cues in both healthy and depressed samples. Williams et al. (1996) used a variant of a specificity instruction task with a neutral valence condition and modified scoring criteria such that specificity was scored on a graded scale from zero (non-specific) to two (specific). Examples of neutral cue words used by Williams included: shop, package, garden, and traveling. Williams et al. reported that a sample who were hospitalized for recent medication overdose as well as psychiatric and healthy control samples reported events with a higher degree of specificity in the neutral, than in a positive or negative valence condition. Despite the uncertainty regarding the importance of valence in memory specificity in depression and inconsistency between individual studies, throughout the memory specificity literature, most studies using specificity instruction tasks have not included a neutral valence condition; so few, in fact, that none of the meta-analyses summarized in the literature review section of this paper analyzed data from neutral conditions of specificity instruction tasks. In absence of a neutral comparison, it is unclear what underlies the typical valence effect in depression. That is, it could be due to impairment/suppression of positive fluency (as is typically interpreted). Alternatively, positive fluency may be on par with neutral, and the finding could reflect enhancement/facilitation of negative fluency relative to baseline. Therefore, although Williams et al.'s study suggests that depressed and healthy samples' reporting on specificity instruction tasks is affected the same way by emotional valence of cues, the inconsistency between the effect of valence on this task and valence in other memory tasks has yet to be explained. More replications are needed in depressed samples to confirm that the relations between emotion and memory specificity are the same between depressed and healthy samples. Importantly, Serrano, Latorre, Gatz, and Montanes (2004) demonstrated that in a 4-week memory specificity intervention targeting positive memories, increased numbers of specific responses to negative, neutral and positive cues were reported following the intervention, but reporting of specific events in response to neutral cues during training was more strongly positively correlated with increase in specific events from pre- to post-intervention than reporting of specific events in response to positive or negative cues. This finding suggests that neutral cues may be more sensitive to treatment effects in depression. In summary, although studies generally suggest that the relation between emotional valence and specificity is the same between healthy and depressed samples, neutral valence conditions may be more sensitive to treatment effects and more research is needed.

Similarly, fluency instruction tasks used in healthy and depressed samples do not include a neutral emotion condition. Based on the MTT Model, it would be predicted that the findings from the memory literature will generalize to future oriented tasks, but given the conflicting findings related to emotion and memory specificity summarized above, it is unclear whether emotional valence might confer an advantage (more events reported) or a disadvantage (fewer events reported) compared to non-emotional events on fluency instruction tasks. This is further complicated in fluency instruction task research because of the Group by Valence interaction (fewer positive events reported by depressed samples). It is unclear whether the valence effect in depression is due to impairment/suppression of positive fluency or enhancement/facilitation of negative fluency relative to baseline. More research is needed to characterize the relationship between valence and fluency of event reporting and how the relationship between emotion and fluency of event reporting is affected in depression.

Studies comparing specificity and fluency instruction tasks support the hypothesis that they assess dissociable sub-processes of MTT event retrieval (D'Argembeau et al., 2010) and that both specificity and fluency instruction tasks are sensitive to cognitive abilities associated with depression (Sarkohi et al., 2011), but several important questions have not been addressed due to methodological issues. Namely, no study has included both past *and* future conditions, along with positive, negative and neutral valence conditions within *each* type of task (specificity and fluency). Moreover, these conditions have never all been examined simultaneously between a sample with depression and a healthy sample. Thus, it is not clear the extent to which reduced specificity across positive and negative valence on AM tasks and reduced positive versus negative event generation on FT tasks in depression are attributable to the differences in task instructions regarding specificity or fluency and/or the differing temporal orientation instructions.

Summary

Despite extensive research examining MTT in depression including recent development of a cognitive training program for depression, there is no consensus regarding the relation between reduced memory specificity and reduced positive fluency in depression. Evidence from healthy and depressed samples has highlighted important distinctions between cognitive processes assessed by specificity and fluency instruction tasks. However, differences between these tasks require systematic examination in depression. In doing so, my dissertation will inform both the avoidance hypothesis of Williams and colleagues' CaR-FA-X model and the generalizability of the MTT model to fluency and specificity tasks in a depressed sample. This dissertation is important and timely in light of proposed causal links between reduced memory specificity and depression and recent interventions designed to increase memory specificity such as mindfulness training (Williams et al., 2000) and MEST (Neshat-Doost et al., 2012; Raes et al., 2009), and the first randomized controlled trial of MEST. Although preliminary findings relating to these interventions are promising, a typical specificity instruction task (the AMT) forms the primary outcome measure. Use of the AMT as a single outcome measure and the CaR-FA-X Model as the basis for an intervention is problematic because the design of the task hinders understanding the nature and cause of the reduction of specific memories reported.

As demonstrated in the literature review above, the CaR-FA-X Model cannot easily explain the lack of valence effects in AM and the inverse-to-hypothesized valence pattern demonstrated on fluency instruction tasks in depression (i.e., reduced positive future fluency). In addition, research supporting the MTT model has shown that MTT is not a unitary process; in particular, Addis and colleagues (2007) have shown that early cognitive processing (Construction) relies on different brain networks than late cognitive processing (Elaboration) in MTT tasks. Similarly, D'Argembeau and colleagues (2010) demonstrated that performance on commonly used MTT tasks, including specificity and fluency instruction tasks relate to different neuropsychological constructs. In depression research, AM and FT findings have received separate attention in the literature, possibly due to a lack of recognition that reduced memory specificity and reduced positive future fluency may have some common neural bases. Importantly, findings from basic cognitive studies (D'Argembeau et al., 2004, 2006) have not been translated to studies of MTT in clinical samples and are not reflected in the CaR-FA-X Model. Because the specificity instruction task and CaR-FA-X Model do not consider the influence of sub-processes as well as moderators and mediators known to be important in MTT, it is possible that some mechanisms of impairment or vulnerability have been overlooked. Based on the MTT Model, the processes underlying AM and FT are the same, and therefore performance differences on MTT tasks should be understood to reflect the processes involved in the task, not temporal orientation of tasks. By determining the degree to which task instructions and temporal orientation instructions account for the differences in performance noted in depression, my dissertation yields insight into the task factors underlying the discrepant valence findings related to reduced memory specificity and reduced positive future fluency. To the extent that different task conditions including task instructions and temporal orientation instructions have different relations to mood and/or depression my dissertation provides important information for advancing understanding of MTT in depression.

In summary, the scope and nature of MTT impairment in depression is unclear. By assessing performance on specificity and fluency instruction tasks in depressed and healthy samples, my dissertation sought to determine the extent to which reduced specificity and reduced positive fluency in depression are linked to effects of task instructions or temporal orientation instructions, and clarify the nature of valence biases in MTT in depression.

Chapter 2: Research Questions

In line with the summary above, my dissertation addressed important research questions regarding three task design factors central to interpreting performance on MTT fluency and specificity tasks in depression: temporal orientation instructions (past and future), task instructions (specificity or fluency) and valence (positive, negative and neutral). The specificity instruction paradigm currently used in depression research (AMT) assesses only past MTT whereas the fluency instruction paradigm (PFT) assesses only future MTT; however, as has been demonstrated throughout this paper, these tasks differ in a number of important ways in addition to their temporal orientation instructions, and the task factors resulting in different patterns of performance on these tasks in depression was unclear.

The MTT Model, developed by Suddendorf and Corballis (1997, 2007), and supported by behavioural (D'Argembeau et al., 2010; D'Argembeau & Van der Linden, 2004, 2006) and brain imaging investigations (Addis et al., 2007, 2008; Addis & Schacter, 2008) suggests that although episodic past and future thoughts vary in intensity on certain factors, they are fundamentally the same in healthy samples. However, the extent of overlap between past and future thinking across a variety of tasks in depression was unclear. The first two research questions I sought to address with my dissertation pertain to the effect of temporal orientation instructions on MTT task performance in depression. *First*, what is the effect of temporal orientation instructions on the number of events reported during a fluency instruction task in depression? *Second*, what is the effect of temporal orientation instructions on the number of specific events reported on a specificity instruction task in depression?

My *third* question about MTT in depression pertained to understanding the effect of emotion on event reporting. The common versions of fluency and specificity instruction tasks do not include a neutral valence condition. The relation between cue valence and specificity of event reporting was unclear since a few studies suggest that the findings of emotional facilitation of memory (D'Argembeau & Van der Linden, 2005; Reisberg & Hertel, 2004) do not generalize to memory specificity studies in depression (Williams et al., 1996). Further, specificity of neutral events may be more sensitive to change over time than emotional events in depressed samples (Serrano, Latorre, Gatz, & Rodriguez, 2004). It was also unclear whether findings of emotional facilitation or suppression reported in previous studies generalize to fluency instruction tasks and how the relation between valence and FT in depressed samples differs from healthy samples. By including neutral valence conditions in specificity and fluency instruction tasks, I hoped to answer the question of whether positive emotional valence facilitates (more total events and/or more specific positive than neutral events reported) or impairs (fewer total events or fewer specific positive than neutral events reported) performance on these tasks and whether the relation between cue valence and event reporting is the same in a depressed sample.

In addition to questions about the role of temporal orientation instructions and cue valence in MTT task performance in depression, my dissertation answered questions regarding the importance of global task focus (specificity or fluency instructions and parameters) in understanding differences in MTT in depression. In particular, my dissertation sought to determine whether differing findings related to valence between specificity and fluency instruction tasks are distinct effects related to specificity or fluency instruction tasks. Therefore, my *fourth* research question was: what is the effect of task instructions on performance in positive and negative valence conditions of fluency and specificity instruction tasks in depression? Recall that on specificity instruction tasks, valence does not emerge as a factor of interest as evident from meta-analyses finding no differences in event specificity between positive and negative conditions in depressed samples (Sumner, Griffith & Mineka, 2010; van Vreeswijk & de Wilde, 2004). In contrast to findings using specificity instruction tasks, on fluency instruction tasks, fewer events are reported in the positive condition relative to the negative condition in depressed samples, suggesting a specific impairment in fluency of positive FTs (for a review, see Szpunar, 2010). By examining general patterns of results (main effects and interactions) and effect sizes between depressed and non-depressed participants on positive, negative and neutral cue valence conditions of specificity instruction and fluency instruction tasks, I aimed to determine whether there are similar or different Group x Valence interactions within the specificity and fluency instruction tasks. In other words, I aimed to determine whether specificity and fluency instruction tasks merely have differing sensitivity to valence effects (i.e., same pattern of valence findings across tasks, different effect sizes) or whether the task factors differing between specificity and fluency instruction tasks are sensitive to distinct MTT heuristic differences in depression (different pattern for valence between tasks). While my questions and hypotheses about temporal orientation and task instruction were framed separately, my analyses assessed whether either factor independently or both are needed to understand the role of valence in MTT in depression.

As summarized in the results section of this paper, examining performance on fluency and specificity instruction tasks clarified which task factors are linked to depressive symptomatology. An additional goal of my study was to conduct correlational analyses guided by main effects and interactions to determine which task conditions are correlated most with depression symptom severity. By determining which task conditions relate to depression, my dissertation permits a more nuanced understanding of MTT in depression than has been possible in the studies summarized above.

Chapter 3: Hypotheses

Based on the literature review, I hypothesized that measures of depression, hopelessness and ruminative tendency would correlate with MTT task performance and that MTT task performance would predict depression symptoms. Based on the CaR-FA-X model, I hypothesized that there would be a significant main effect of Group for the specificity instruction task such that participants with depression would have lower scores in the Past conditions than participants without depression. Based on the FT literature, I hypothesized that there would be an interaction between Group and Valence on the fluency instruction task such that the depressed sample would report fewer events in the Positive than Negative Future condition. Based on the MTT Model, I hypothesized that the fluency instruction task and specificity instruction task would demonstrate significant positive correlations between all conditions within each task and all conditions between tasks, with magnitude of correlations being higher within than between each task, and within temporal direction and valence conditions than between. Lastly, I hypothesized based on the MTT Model that patterns related to Group and Valence would be mirrored across both Temporal Directions. Based on these hypotheses, I predicted that the same patterns would be displayed in correlations of conditions within the fluency instruction task, specificity instruction task, and correlations between conditions of both tasks. Specifically, all conditions of each task would be positively correlated. Past conditions would correlate more strongly with other Past Conditions than with Future conditions and Future conditions would correlate more strongly with other Future conditions than with Past conditions. Similarly, Valence conditions (Negative, Neutral, and Positive) would correlate more strongly with corresponding Valence conditions in the alternative temporal direction than with alternative Valence conditions. For instance, I predicted that the Past Positive condition of the fluency instruction task would correlate more strongly with the Future Positive condition of the fluency instruction task than either the Future Neutral or Future Negative conditions of the fluency instruction task. The patterns outlined in the example above would also be displayed in the specificity instruction task. When correlating conditions of the fluency and specificity instruction tasks, a similar pattern of stronger positive correlations within matching Temporal Direction and Valence conditions than between would also be evident.

Chapter 4: Method

To address the questions outlined in the previous section, I quantitatively examined the effects of temporal orientation instructions and cue valence within fluency and specificity instruction tasks and qualitatively examined the effects of temporal orientation instructions and cue valence between fluency and specificity instruction tasks in depressed and healthy samples using a four-way mixed within (MTT task factors: task instructions, temporal orientation instructions, cue valence) and between subjects (Groups: currently depressed and never depressed) design. A goal of the study was to clarify existing patterns in MTT task performance in depression. More specifically, I aimed to replicate typical discrepancies reported between depressed and non-depressed samples in number of positive and negative events reported on future-oriented fluency instruction tasks and number of specific events reported on past-oriented specificity instruction tasks. Therefore, my methodology was based on methods used in previous and ongoing research using MTT fluency and specificity instruction tasks in depression.

The following sections outline central methodology pertaining to sample definition, identification and recruitment, characterization of psychological and cognitive factors demonstrated in previous studies to be useful to interpret performance on fluency and specificity instruction tasks, and parameters of the specificity and fluency instruction tasks used in this study. Although the focus of my dissertation was to address the four research questions articulated in the previous section, to the extent possible without unduly lengthening my study protocol, or compromising my study design and feasibility, I included a small number of measures and tasks assessing constructs of secondary interest that are linked to the MTT and CaR-FA-X Models. The purpose of including these additional assessment measures, which will be described later, was to rule out potential study confounds, confirm generalizability and replication of the results, and help to contextualize my findings within the CaR-FA-X and MTT Models. I have briefly outlined a theoretical and practical rationale for each of the measures I included following an overview of my core methodology.

Procedure

Participants were recruited during the Fall 2014 and Winter 2015 academic terms. Participants were recruited through the undergraduate research participant pool (Sona) at Ryerson University. Prospective participants completed an online prescreening questionnaire package and those fulfilling eligibility criteria outlined in the section to follow were given the option to participate in the study. Each participant was tested individually. The study took place during one two-hour lab session, during which participants completed MTT specificity and fluency instruction tasks, self-report measures of depression symptom severity, a semi-structured diagnostic interview, and questionnaires and cognitive tasks assessing additional constructs linked to MTT and the CaR-FA-X Model in previous studies. The sample eligibility criteria and assessment measures are described below.

Sample Definition and Assessment of Depression

Methods to select and categorize participants and assess depression were chosen with two aims in mind: 1) to maximize the odds of replicating main findings in the literature regarding MTT specificity and fluency in depression; 2) to be generalizable to ongoing clinical research on treatment of MTT specificity (Medical Research Council, 2000). Research on MTT fluency and depression is limited and has mostly been undertaken by a single network of collaborative researchers (including MacLeod and Matthews), resulting in relatively homogenous methods regarding identification and characterization of depressed participants. Studies conducted by this group often used convenience samples recruited from medical clinics on the basis of an existing diagnosis of depression or recent self-harm behaviour (for example: MacLeod et al., 1993, 1997; MacLeod & Salaminiou, 2001); depressed student samples have also been used (MacLeod & Byrne, 1996). The measure used to characterize depression severity most consistently within this literature is the revised version of the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996; Steer, Cavalieri, Leonard, & Beck, 1999). However, the link between MTT specificity and depression has been examined using a greater variety of methods, the implications of which have been scrutinized. Although reduced AM specificity is hypothesized to be a stable trait in people with depression and has been argued to be linked to depressive symptoms rather than diagnosis (Dalgliesh et al., 2007), meta-analyses by van Vreeswijk and de Wilde (2004) and Sumner et al. (2010), have found that clinical diagnosis of depression moderates the relation between depression symptom severity and performance on memory specificity instruction tasks. An item-response analysis of the AMT (Griffith et al., 2009) and a review of psychometric issues in Overgeneral AM research by Griffith et al. (2012) similarly highlight the importance of sample definition in interpreting performance on memory specificity tasks in depression, each noting that the standard specificity instructions are sensitive to relatively extreme differences between groups but less sensitive to detecting more subtle associations than versions of the specificity task using different task instructions (Debeer et al., 2009; Raes et al., 2007). Based on these findings, although people with depressive symptoms who do not meet diagnostic criteria for depression as well as people with remitted depression demonstrate similar patterns to those with current depression, the magnitude of the relation may be smaller, reflecting both the trait and state qualities of low event specificity in depression. To maximize the chances of replicating findings from the literature, populations of interest and measures to identify and characterize depression were chosen with reference to moderators of memory specificity and depression identified in the above review papers. Therefore, groups for this study consisted of participants with current or past depression based on DSM-IV criteria (depressed group) and depression severity scores ranging from mild to severe, and participants who have no history of depression (never depressed group). In the paragraphs to follow, descriptions and rationales for each of the measures used to identify participants with depression, assess full diagnostic criteria for depression, and assess depression severity are included, followed by a description of how each measure was used to assess and characterize participants in my study.

As indicated in the procedure section, prior to participating in the study, prospective participants completed a study eligibility assessment via computerized prescreen including the two-item version of the Patient Health Questionnaire (PHQ-2; Kroenke, Spitzer, & Williams, 2003) and two items from version 6.0 of the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). The PHQ-2 was used to identify prospective participants for the depressed group and the MINI was used both to identify prospective participants for the never depressed group and confirm group eligibility, as described below.

The PHQ-2 was developed for primary care settings and has been validated as a screening tool for depression using a large sample. In the PHQ-2, participants are asked how often they have been bothered by anhedonia and low mood during the previous two weeks; participants indicate their response for each item by selecting one of four options ranging from ‘Not at all’ to ‘Nearly every day’. Thus, the more frequently a participant reports that symptoms have occurred, the higher their score will be. The PHQ-2 was used to identify participants likely to meet diagnostic criteria for depression based on measures administered during the study session. Kroenke et al. (2003) reported that among adults in a primary care setting, scores on the

PHQ-2 of three or greater out of a possible six had sensitivity of 83% and specificity of 92% using diagnoses based on structured interviews conducted by a mental health professional as the criterion. Therefore, a cutoff score of 3 was used to identify prospective participants for the depressed group. Diagnostic status and group eligibility were confirmed using measures described later in this section completed during the study session.

In addition to the PHQ-2, study eligibility was assessed using two screening questions from the DSM-IV Mood Disorder Module of the MINI (Sheehan et al., 1998). Diagnostic criteria for depression (and other forms of psychopathology) were also assessed in both groups using the MINI, a brief structured interview designed for use as a research measure. This measure was chosen to determine group eligibility and characterize psychopathology in this study because it has good reliability and validity for depression, is widely used, was designed for research studies (valid for use in general populations and clinical populations), can be administered quickly and easily, and screens for presence of common current and past disorders including other mood disorders, anxiety disorders, substance abuse and dependence disorders, and psychotic disorders. Although the MINI 6.0 was developed and validated for DSM-IV-TR, diagnoses and reliability data regarding DSM-5 diagnoses had not been published at the time of data collection, so this was the most up to date measure available to use. Items concerning Major Depressive Episodes (MDE; American Psychiatric Association, 2013) as well as most other conditions assessed by the MINI version used in my dissertation correspond closely to DSM-5 criteria because the criteria have changed only minimally (American Psychiatric Association, 2013). During the pre-study assessment, participants were asked to answer MINI questions A1a: “Were you ever depressed or down, most of the day, nearly every day, for two weeks?” and A2a: “Were you ever much less interested in most things or much less able to enjoy the things you used to enjoy most of the time, for two weeks?”), that assess for positive history of DSM-5 criteria A. 1 and A. 2 for MDE (one of which is required for diagnosis). These MINI questions on the prescreen were used to identify prospective participants who were unlikely to have had MDEs in the past (never depressed group) prior to being invited to participate in the study.

Participants who responded ‘no’ to both MINI items and score two or lower on the PHQ-2 were eligible to participate in the Never Depressed Group (NDG). Participants with PHQ-2 scores of three or above were eligible to participate in the Depressed Group (DG). The depressed sample in this study were not required to have had a history of depressive episodes aside from the current episode; therefore, participants with PHQ-2 scores of three or higher were eligible to participate in the study regardless of response to the MINI screener questions.

Both groups of participants completed the full MINI as well as the BDI-II during the study to confirm eligibility based on current or historical diagnostic status of depression and symptom severity at the time of testing. The BDI-II has been demonstrated to be a reliable and valid measure of depression symptom severity in diverse populations, including depressed and other psychiatric samples as well as community and undergraduate samples (Dozois & Covin, 2004; Storch, Roberti, & Roth, 2004; Whisman, Perez, & Ramel, 2000). The BDI-II has been used to identify and characterize depressed samples in previous MTT research, including recent trials of the MEST protocol (Medical Research Council, 2000; Raes et al., 2009). In this study, BDI-II scores were used as a secondary criterion to identify participants with a current MDE as well as characterize depression severity in the sample. In order to be eligible for the NDG, participants must have scored below 10 on the BDI-II (Kendall, Hollon, Beck, Hammen, & Ingram, 1987) and not meet criteria for a current or past MDE based on their responses to the MINI. In order to be included in the DG, participants must have scored above 13 on the BDI-II, indicative of at least mild depression based on previous studies and meet diagnostic criteria for current or past MDE based on the MINI.

Participants in either group who reported symptoms consistent with disorders other than a mood disorder based on responses to the MINI were not excluded from the study because it is unclear whether history of comorbidities affects MTT. However, comorbidities are reported in the next section.

Participants who did not meet eligibility criteria for either group at the time of testing (i.e. fully remitted depression, subclinical depression) participated in the study but were not counted toward the total number of participants in either the NDG or DG. These data were not included in the main analyses reported. Means (*M*) and standard deviations (*SD*) for most independent and dependent study variables per group based on initial inclusion and exclusion criteria are summarized in Appendix A. Data from excluded participants were of interest in their own right, since reduced positive future fluency and Overgeneral AM have been shown in a variety of samples with depression (Griffith et al., 2012; Sumner et al., 2010). Further, it was unclear from the study planning phase how many participants would meet the a priori inclusion criteria for the DG and how inclusion criteria might affect study group composition because this was an undergraduate student sample. Data from excluded participants were examined using exploratory analyses to understand their possible relevance to the DG and NDG results in my dissertation. Although full exploration of the significance of these data was beyond the intended scope of my dissertation, more research is needed to understand the potential influence of intermediate mood diagnoses and symptoms on MTT.

Power Analysis

Power analyses were conducted prior to the start of data collection using G*Power 3.0.10 (Faul, Erdfelder, Lang, & Buchner, 2007). Studies using MTT tasks in healthy samples and fluency instruction tasks with depressed samples often report highly significant effects with small samples. For example, D'Argembeau & Van der Linden (2004) were able to detect interaction effects between temporal orientation and valence conditions using a sample of 40 healthy adults and MacLeod et al. (1996) reported a significant four-way interaction between group, temporal orientation, temporal distance from present, and valence with a total sample of 51 including anxious, depressed and healthy samples. On the other hand, a study by D'Argembeau and colleagues (2006) reported that the difference between performance on past and future conditions of a fluency instruction task in a healthy sample of 100 participants was small (d = 0.29) but significant. Studies of MTT specificity in depression typically have medium or larger effect sizes related to group, although non-significant effects are also reported (see studies reviewed in Table 1 by Williams et al., 2007). Due to the uncertainty of predicting effect size a priori given the number of factors in the tasks I used and variability of methodology and results reported in previous studies, as well as practical considerations, I estimated the sample size required to detect a large effect (Cohen's d = 0.8) on an independent samples t-test. Between-subjects tests have less power to detect effects, therefore I tailored my assessment of estimated sample size to the independent samples t-test to obtain a conservative estimate in this regard. Incorporating common parameters for power analyses (α = .05, β = .20), I calculated that 52 participants (26 per group) would be sufficient to detect a large effect.

Participants

Participants in the NDG (*N*=27, 22 female) did not endorse current or past depression (BDI-II total score < 10, suicidality item = 0; no mood disorders endorsed on MINI), other mental health conditions on the MINI, current significant anxiety or hopelessness (BAI total score < 21, BHS total score < 3), other conditions or characteristics known to affect performance on cognitive tasks (i.e. neurological conditions, recent drug use) and scored above a recommended cutoff score of four for detecting low motivation on the Digit Span test (Babikian, Boone, Lu & Arnold, 2006). Participants who reported diagnoses of learning disability (*N* = 6), prior concussion (*N* = 1), neurological disorders (*N* = 2), current use of prescription drugs for a mental health condition (*N* = 2) were excluded from the NDG; no participants who met criteria for the NDG reported recent recreational drug use. In the total sample, only four participants scored below five on the Digit Span Test, including one participant in the DG; this participant's score was not low on any other measures, nor were they an outlier, therefore their data were included in the analyses.

Participants in the DG (*N*=44, 35 female) scored above 13 on the BDI-II and reported either current or a past mood disorder on the MINI. Participant characteristics for the DG and NDG are summarized in Table 1. T-tests revealed that as predicted, the DG scored higher than the NDG on BDI-II and measures of other clinical constructs including anxiety, hopelessness, ruminative tendency, and number of potentially traumatic events. Further information about clinical measures administered in my dissertation is included in the Secondary Measures section, beginning on page 56.

Of participants in the DG, 28 met criteria for recurrent depression, meaning that they have had at least 2 previous MDEs. Thirty-seven percent of participants in the DG who answered

Table 1: Participant characteristics by group and t-tests

	Group	*N*	*M*	*SD*	*t*	*p*
Participant Age	NDG	27	20.07	3.93	0.11	.914
	DG	43	19.95	4.85		
Beck Depression Inventory Score	NDG	27	3.52	2.61	-13.21	<.001
	DG	44	22.91	7.33		
Beck Anxiety Inventory Score	NDG	27	5.70	3.31	-8.67	<.001
	DG	42	22.26	9.54		
Beck Hopelessness Scale Score	NDG	27	1.48	1.16	-3.38	.001
	DG	38	7.11	4.88		
Ruminative Response Scale Score	NDG	27	31.67	8.41	-6.20	<.001
	DG	38	49.61	13.25		
Trauma History Screen Total Events Reported	NDG	27	1.48	1.12	-2.16	.035
	DG	38	8.92	17.88		
Controlled Oral Word Association Test (FAS)	NDG	27	35.30	8.09	-0.80	.429
	DG	42	37.17	10.35		
Categorical Fluency (Animals)	NDG	27	21.37	4.34	0.08	.939
	DG	42	21.29	4.58		
Longest Digit Span Forward Score	NDG	27	6.30	1.07	-1.22	.225
	DG	40	6.70	1.47		
Digit Span Forward Score	NDG	27	9.33	2.43	-1.37	.177
	DG	39	10.13	2.25		
WRAT Reading	NDG	27	58.74	5.40	-1.70	.094
	DG	39	60.77	4.28		

Note: all scores reported are raw scores; varied Ns are due to missing data. DG = Depressed Group; NDG = Never Depressed Group, WRAT = Wide Range Achievement Test.

the screening question reported a non-English primary language compared to fifteen percent in the NDG. Two participants in the DG reported having been diagnosed with a learning disability, one reported having had a prior concussion, and one reported having been diagnosed with a neurological disorder. Eight participants in the DG reported currently taking prescription medication for a mental health condition and five reported that they had recently used recreational substances. These participant characteristics were noted, but they were not excluded from the analyses. Twenty-five participants endorsed symptoms of at least one comorbid disorder. Current comorbidities in the DG included Generalized Anxiety Disorder (*N* = 4), Obsessive Compulsive Disorder (*N* = 3), Post Traumatic Stress Disorder (*N* = 2), Alcohol Abuse or Dependence (*N* = 2), Substance Abuse or Dependence (*N* = 2), Panic Disorder or Agoraphobia (*N* = 6), Psychotic Disorder (*N* = 3), Eating Disorder (*N* = 2), and Social Anxiety Disorder (*N* = 2).

MTT Tasks

In addition to standard versions of the well-established past and future-oriented specificity and fluency instruction tasks, my project is innovative in that I expanded on previous studies to include conditions to assess both temporal orientations (past and future) using specificity *and* fluency instructions similar to what has been done using healthy samples (D'Argembeau et al., 2010). This within-subjects design allowed me to assess MTT specificity and fluency among depressed individuals in both past and future conditions. Within each task, I also assessed performance under positive, negative and neutral valence conditions. Throughout the remainder of my dissertation, I use acronyms to distinguish the Fluency Instruction Task (FIT) and Specificity Instruction Task (SIT) used for my dissertation from other fluency and specificity instruction tasks in the literature, where I spell out these terms. Participants completed these tasks in a single session before any other study tasks to reduce mood priming effects and with fluency/specificity task instructions, temporal orientation instructions and cue valence conditions counterbalanced across participants. MTT tasks were audio recorded and transcribed to facilitate reliable scoring.

Specificity Instructions Task. All participants completed a SIT modeled after the instructions of the AMT (Williams & Broadbent, 1986). During the AMT, participants are instructed to retrieve a memory from their personal past in response to a cue word. The original AMT stimuli consist of five pleasant words (happy, safe, interested, successful, and surprised) and five unpleasant words (sorry, angry, clumsy, hurt, and lonely) that have been previously demonstrated to reliably sample positively and negatively valenced AM in a healthy sample (Williams & Broadbent, 1986). Participants are told that the memory can be something that happened recently or a long time ago and that it can be an important or trivial event but should be of something that happened at a particular place on a particular day; participants are told that they may not repeat events across trials or report events that occurred within the past week. Participants are given 60 seconds to respond to each word. Words from the pleasant and unpleasant lists are presented in alternating order and latencies until the participant responds are recorded. In subsequent studies using the AMT, task parameters have been modified as appropriate to the goals of the study including cue words used, the length of time to respond, wording of instructions, presentation format (written, oral or both), inclusion of practice trials, provision of examples of acceptable and unacceptable responses, and presentation of cues in sentence form (Griffith et al., 2012).

Task instructions for the SIT can be found in Appendix C. To create the SIT, I adapted the AMT instructions to include past and future conditions and neutral cues in addition to positive and negative cues. Instruction wording was adapted to be as similar as possible for both past and future conditions: “Tell me one specific moment or event from your past that the word X reminds you of” and “Tell me one specific moment or event from your future that the word X reminds you of”. Stimuli (10 positive, 10 negative and 10 neutral cue words) were selected from cues used in the original AMT (Williams & Broadbent, 1986) and additional cue words, including neutral words, used in the AMT in subsequent studies (Williams et al., 1996; Park, Goodyer, & Teasdale, 2002) to create two matched 15-item lists. Psycholinguistic properties of the lists were statistically compared prior to data collection using the Warriner and MRC psycholinguistic databases (Warriner, Kuperman, & Brysbaert, 2013; Wilson, 1987). For consistency between Lists, Valence conditions, and to be consistent with normed datasets, seven of the cue words used were slightly modified before administration ("surprised" to "surprise", "successful" to "success"). The word Lists and Valence categories were statistically compared on several characteristics available in the MRC and Warriner databases, including frequency, concreteness, familiarity, imageability, number of syllables, number of phonemes, valence and arousal. Nonparametric statistics were used because the independent variable (Valence) was ordinal and data were not normally distributed. The Jonckheere-Terpstra Test indicated that there were no significant differences between the Lists or Valence conditions on any of the characteristics examined aside from predicted difference in word valence between all Valence conditions ($T_{JT} = 300.00$, $z = 5.704$, $p < 0.001$); in fact, the Valence Rating distributions were non-overlapping and support that the word classification based on Valence was valid. Descriptive statistics for the stimuli by Valence can be found in Table 2.

As recommended by van Vreeswijk and de Wilde (2004) and Griffith et al. (2012), I used audio-visual presentation of cues and audio-recorded responses to increase scoring reliability. Trial duration has been reported to be a significant moderator of memory specificity (van Vreeswijk & de Wilde, 2004), with largest effects between 30s and 60s, but Raes et al. (2007) created a modified version of the AMT for use among healthy undergraduate students because a typical version proved too easy during piloting. Dr. Caitlin Hitchcock (2014), a collaborator of Dr. Mark Williams, noted in a personal communication that their research team has adopted a 30s trial duration to increase task difficulty. Therefore, participants in the current study

Table 2: Descriptive statistics of SIT stimuli by valence

		N	*M*	*SE*	*Minimum*	*Maximum*
Valence Rating	NEG	10	2.69	0.16	2.10	3.56
	NEU	10	5.94	0.23	4.63	6.77
	POS	10	7.60	0.13	7.00	8.47
	Total	30	5.41	0.39	2.10	8.47
Arousal Rating	NEG	10	4.92	0.21	3.49	5.84
	NEU	10	4.00	0.29	2.52	5.25
	POS	10	5.44	0.33	3.14	6.62
	Total	30	4.79	0.19	2.52	6.62
Familiarity Rating	NEG	10	558.30	7.92	505	589
	NEU	10	562.60	13.13	497	625
	POS	10	568.80	10.18	511	621
	Total	30	563.23	5.97	497	625
Imageability Rating	NEG	10	444.50	19.08	356	550
	NEU	10	464.00	32.73	305	587
	POS	10	492.10	19.96	406	615
	Total	30	466.87	14.24	305	615
Meaningfulness (Colorado Norms)	NEG	8	481.13	13.66	421	530
	NEU	7	465.43	13.23	427	518
	POS	8	499.75	15.03	445	568
	Total	23	482.83	8.30	421	568
Brown Verbal Frequency	NEG	7	8.86	4.59	1	36
	NEU	8	18.63	6.40	3	55
	POS	10	6.90	2.93	1	32
	Total	25	11.20	2.77	1	55
Thorndike-Lorge Written Frequency	NEG	10	625.50	261.47	73	2859
	NEU	10	763.70	246.47	24	2299
	POS	10	892.40	212.59	95	2143
	Total	30	760.53	135.8	24	2859
Kucera-Francis Frequency Rating	NEG	10	52.20	16.07	10	179
	NEU	10	81.90	28.69	9	330
	POS	10	53.30	8.99	3	98
	Total	30	62.47	11.26	3	330
Kucera-Francis Number of Categories	NEG	10	10.90	0.82	7	15
	NEU	10	11.80	1.06	4	15
	POS	10	11.30	1.03	3	14
	Total	30	11.33	0.55	3	15
Concreteness Rating	NEG	10	334.70	21.73	262	504
	NEU	10	426.10	42.4	238	580
	POS	10	386.30	26.65	295	533

	Total	30	382.37	18.88	238	580
Number of Letters	NEG	10	5.00	0.42	3	7
	NEU	10	6.50	0.54	4	9
	POS	10	5.90	0.61	4	10
	Total	30	5.80	0.32	3	10
Number of Phonemes	NEG	10	3.80	0.33	2	5
	NEU	10	5.30	0.54	3	8
	POS	10	4.90	0.66	3	10
	Total	30	4.67	0.32	2	10
Number of Syllables	NEG	10	1.30	0.15	1	2
	NEU	10	2.10	0.23	1	3
	POS	10	1.60	0.22	1	3
	Total	30	1.67	0.13	1	3

Note: *N*s< 10 reflect lack of psycholinguistic normative data for words in that condition on that factor.

were permitted 60s per trial, but scoring was based on the first response provided within 30s to increase difficulty and permit re-scoring as a possible follow-up.

Each response was categorized using standardized scoring criteria guided by documentation provided by several experts in the field via personal communication (Crane, 2015; Hitchcock, 2014; Williams, 2014, 2015) as either specific (i.e., lasting a day or less and taking place at a particular location), categorical (a type of event that occurs repeatedly without a specific time or place reported), extended (an event with a duration > 1 day), semantic association (verbalization of a non-event), or omission (no response). Including initial instructions, completion of at least two correct practice trials and transitions between trials, the SIT took approximately 35 minutes to administer. An administration protocol for the SIT can be found in Appendix B.

Fluency Instructions Task. I derived instructions for the FIT from those of the Future Fluency Task (FFT) developed by MacLeod and colleagues (1993; 1996); a full description of the task adapted from MacLeod and colleagues including administration script can be found in Appendix B The FFT consists of 6 trials that vary according to valence (looking forward to or not looking forward to) and temporal distance (1 week, 1 year, five to ten years). For each combination of valence and temporal distance, participants are instructed to list as many things as they can within one minute. In addition to these trials, the FIT included an equal number of past-oriented trials, similar to what was done by D'Argembeau et al. (2010). In addition, a neutral condition that was developed for a previous study in our lab (Luong, Burley, & Girard, 2014) was administered. Thus, the FIT consisted of 18 trials of one-minute duration each, that vary by temporal orientation (past, future), valence (positive, negative, neutral) and temporal distance (next week, next year, next five to ten years). I audio-recorded responses to the FIT to facilitate scoring; in line with methodology used by MacLeod and colleagues (MacLeod, March 2015 via personal communication) I summed the total number of items per 60-s trial. Including the initial task instructions and transitions between trials, administration time for the FIT was approximately 20 minutes. An administration protocol for the FIT can be found in Appendix B.

Secondary Measures

In addition to the MTT tasks and depression assessment measures, participants answered demographic questions and completed several measures assessing constructs aside from depression that have been implicated in performance on MTT tasks in previous studies. These measures were used to assess potential study confounds, and link my findings to other important constructs and findings related to the MTT and CaR-FA-X Models. For each measure, I have provided a brief rationale and summary of findings from previous studies. The measures outlined below were administered in the order in which they are described in this section.

To assess whether the samples differed in general cognitive functioning and ensure comparability with samples in previously published studies, I administered measures of cognitive functioning including academic achievement, general cognitive ability (IQ), and executive functioning. To characterize executive functioning in my samples, I administered the Controlled Oral Word Association Test (FAS; Lezak, 2004; Tombaugh, Kozak, & Rees, 1999) and Category Fluency Test (Animals). These two classic neuropsychological tests assess phonemic and semantic fluency, respectively; due to the similarities between the structure of these tasks and fluency instruction tasks, they are commonly administered to samples in studies of MTT and have been associated with MTT performance in healthy samples and depressed samples (Dalgliesh et al., 2007; D'Argembeau et al., 2010). In addition, I included the Digit Span Forward (DSF) subtest of the Wechsler Adult Intelligence Scales, Fourth Edition (WAIS-IV; Wechsler, 2008) to characterize working memory ability and assess participant effort (Babikian et al., 2006); performance on the Digit Span Forward task has previously been linked to performance on specificity instruction tasks (Dalgliesh et al., 2007). Lastly, I administered the Reading subtest of the Fourth Edition of the Wide Range Achievement Test (WRAT-4; Wilkinson & Robertson, 2006) as an index of general cognitive functioning and academic achievement; this measure correlates significantly with IQ score on the WAIS-IV and was chosen over other indices of general cognitive functioning due to the centrality of verbal tasks in this study.

In addition to collecting demographic information and administering cognitive tasks to characterize the samples, several other measures were administered to address potential study confounds, and examine hypothesized relations between MTT task variables and other clinical and individual difference constructs linked to MTT and depression. Constructs linked to MTT task performance in previous studies that I assessed include current anxiety symptoms, hopelessness, ruminative tendencies and history of trauma. The measures used to assess these constructs, all of which are described in the following paragraphs, have acceptable reliability for use in research studies and have been used in student and community samples.

MacLeod and colleagues (1997) reported that anxiety, as measured using the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988), was correlated positively with number of negative FTs reported. Given that my third research question pertained to the effect of cue valence on event reporting and that anxiety has been associated uniquely with negative FTs in previous studies, I included the BAI in my study protocol. The BAI has also been used in other studies of MTT in depression including the ongoing RCT for MEST (Medical Research Council, 2000; Sarkohi et al., 2011). Thus, participants completed the BAI to enable examination of the potential relation between anxiety symptoms and performance on MTT fluency and specificity instruction tasks.

The first studies of MTT specificity and fluency were conducted in samples who recently attempted suicide (Williams & Broadbent, 1986; MacLeod, Rose, & Williams, 1993). Although the findings have since been replicated in samples with depression who have not attempted suicide, the relation between depression, hopelessness and MTT task performance is still unclear and therefore hopelessness and suicidality are often assessed in studies of MTT and depression. Recent studies (Medical Research Council, 2000; Raes et al., 2009) have used the Beck Hopelessness Scale (BHS; Beck & Steer, 1988; Beck, Weissman, Lester, & Trexler, 1974), a 20-item self-report measure to examine changes in hopelessness as a potential mediator of improvement in depression in treatment studies. Although hopelessness and depression are typically correlated, it is unclear whether hopelessness or depression symptom measures will be more closely linked to MTT task performance; therefore I included the BHS to characterize relations between MTT task scores and hopelessness.

As summarized by Williams et al. (2007), MTT specificity has also been linked to ruminative tendencies: Consistent with the CaR-FA-X Model, when induced to ruminate, depressed, but not healthy samples, demonstrate lower MTT specificity and worse mood. Lavender and Watkins (2004) examined the effect of rumination on MTT fluency by administering a fluency instruction task following induced rumination or distraction in a depressed sample; contrary to what was predicted based on MTT specificity studies, rumination resulted in increased negative and positive FTs. Although my dissertation is not designed to address questions relating to the rumination component of the CaR-FA-X Model, to my knowledge, no studies to date have assessed the relation between rumination and MTT fluency and specificity instruction task performance using a within-subjects design. In addition to having distinct relations with depression, performance on MTT specificity and fluency instruction tasks may also relate differently to rumination. Therefore, I administered the Ruminative Responses Scale (RRS; Treynor, Gonzalez, & Nolen-Hoeksema, 2003), a 22-item self-report scale that measures the trait-like tendency to respond to distress by passively ruminating about problems or past situations (Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008).

Lastly, history of trauma is also associated with reduced memory specificity (e.g. Wessel et al., 2002). Although the MINI can detect current Post Traumatic Stress Disorder (PTSD), it does not assess for history of traumatic experiences or history of PTSD and this has been noted as a common confound in research on the CaR-FA-X Model (Griffith et al., 2012; Sumner, 2012; Williams et al., 2007). To facilitate discrimination of the effects of depression from the potential effects of trauma history on memory specificity, I administered the Trauma History Screen (THS; Carlson et al., 2011). The THS is a 14-item self-report measure that assesses number, age and frequency of occurrence of events that may satisfy criterion A of the DSM diagnostic criteria for PTSD in approximately 8 minutes. Following identification of possible traumatic events, the test taker completes additional questions to characterize the event and clarify whether the event would likely satisfy criterion A of the PTSD diagnostic criteria. The THS has demonstrated reliability and validity in clinical, community and undergraduate samples and is correlated with PTSD diagnosis.

MTT Task Scoring Reliability

Scoring and administration guidelines for the FIT and SIT were obtained from descriptions in published studies and through personal communications with Dr. Andrew MacLeod (March, 2014), Dr. Mark Williams (October, 2014), Dr. Catherine Crane (April, 2015), a collaborator of Williams, and Dr. Caitlin Hitchcock (April, 2015), a post doctoral student working with Dr. Tim Dalgliesh. Scoring documentation for the FIT and SIT can be found in Appendices A through C.

Each participant's MTT responses were audio-recorded for later scoring, with the exception of one participant who did not provide consent to be audio recorded. The MTT tasks took approximately one hour to administer, and therefore required at least an additional hour of scoring to review the audio-recording for each participant. To assist with this task, two post-bachelor's level research assistants were recruited and trained for over two months in the scoring of these tasks.

The research assistants were provided with descriptions of scoring for each task from the literature as well as documentation and clarification from the task creators obtained through email, as outlined above. Although the MTT literature and task developers provided guidelines regarding scoring, in practice, scoring was complex and there were unforeseen scoring ambiguities within each task. These challenges have been outlined in the subsections to follow.

While scoring, additional rules of thumb for making decisions in scoring the FIT and SIT were agreed upon by myself, research assistants and the researchers; these additional scoring rules can be found in Appendix D. The research assistants were naïve to study hypotheses and were unfamiliar with the literature on MTT and depression.

The research assistants were provided with feedback from me on their scoring decisions for the MTT task data of five consecutive participants arbitrarily chosen and instructed to score at least five additional participants independently, at which point the reliability of scoring was examined by comparing the research assistants' scores with my scores; the results of reliability analyses for each task are reported below.

FIT. Material provided by MacLeod regarding the FIT indicated that to score it, one must add up "the total number of responses given in a particular condition time period […]. Where a subject repeats a response across different time categories only include it the first time it is mentioned" (MacLeod, personal communication, 2014). While scoring FIT data from my study, types of ambiguities not mentioned in the literature that were noted included how to count events when participants reported a recurring event (repeats) in several different temporal conditions (e.g. "walking the dog"), or across different valence conditions (e.g. "moving to Toronto"). In addition, participants would occasionally report several events in summary form as a single statement (e.g. "I went to three cottages last summer") or two seemingly different events in a single statement ("Fighting with my parents…and my brother"). Dr. MacLeod indicated that in his studies, for the first two examples outlined above, neither would be counted as a repeat; for the third and fourth examples, these would be counted as single events. General scoring principles that were used to help clarify ambiguous participant responses included giving participants the benefit of the doubt if it is unclear whether a response might be a repeat, and counting events using phrasing and sentence structure rather than inferring number of events based on content, since this could be too inferential and could introduce rater-based bias.

For the FIT, the dependent variable was the total number of events reported in each condition. Between myself and the first research assistant, there was a high correlation between the total number of events we each counted across all conditions for nine research participants ($r = 0.99$). The second research assistant scored only five participants and our correlation was small in magnitude ($r = 0.26$). On the basis of these correlations and observations during training, I concluded that the second research assistant did not have a strong understanding of the construct being measured. Therefore, the first research assistant contributed to scoring the FIT whereas the second research assistant did not. To examine the reliability of scoring, intraclass correlation coefficients (ICC) and mean difference scores were computed for nine consecutive participants arbitrarily selected from the midpoint of the data collection period. A two-way mixed effects model of absolute agreement (ICC model 1,A; McGraw & Wong, 1996) was used as a conservative measure of reliability because it takes into account the correlation between the raters' scores as well as the mean difference. Nonetheless, the absolute mean difference between raters' scores was |0.889| and not significant, $p = .937$. The ICC between the research assistant and I was .996 (95% *CI*: .984-.999), $p < .001$. These indicators support very high interrater reliability for the FIT scores in this study.

Of the 71 participants who met criteria for inclusion in the DG and NDG, the first research assistant scored and entered data for 21. There were no significant differences between the mean number of events reported on the FIT for participants whose data I scored compared to participants whose data were scored by the research assistant ($t = -1.16$, $p = .250$).

SIT. For the SIT, Williams provided documentation with the description that a response should be counted as specific if it is “an event lasting a day or less, which occurred at a certain place and time” and “if the participant arrives at the memory on their own without prompting” (Williams, personal communication, 2014). Examples of specific events provided in documentation shared by Williams and Dalgliesh include “the birth of my daughter,” and “the day I got my exam results.” Examples of generic responses included “mountain climbing in Wales” and “the lead up to me being made redundant [getting fired].” Instructions from Williams (2015) for the AMT and instructions from Dalgliesh (Hitchcock, personal communication, 2014) regarding the AMT task used in the MEST included the additional provisions that prompting could be used if the initial response was ambiguous to clarify whether the event described was likely specific and that participants’ first responses only should be considered. Because prompting was intended only to clarify whether an initial response was specific, if a participant did not respond to a prompt within 10 seconds, it was inferred that any information they reported was generated in response to the prompt rather than the cue word and the response was not used to clarify event categorization.

Occasionally during the SIT, a participant would give an invalid response: for instance, an event from the past in the future condition, or a response that was within 7 days of the study date. Participants who gave invalid responses were given a corrective prompt, and their corrected response was scored. Some participants provided the same event for multiple cues; in this case, participants were prompted with the instruction “can you think of a different event?” but their response was scored as a repeat. During the main analyses, repeats were considered as a distinct category of non-specific response, since for those trials, the participant was unable to retrieve a unique specific event for that particular prompt. For events that reoccur, such as birthdays, the event was only scored as specific if when describing the event, the participant noted some type of detail about the event that distinguished it from other iterations. For instance, if a participant reported “my brother’s birthday next year; we always go to the same restaurant, and my whole family will be there as usual,” this would only be counted as specific if the participant described something unique about the event, such as “I am picturing a chocolate birthday cake with ‘22’ on it.”

The two research assistants, Dr. Girard and I agreed that future events seemed less rich in detail and more generic than past events and agreed that future events were subjectively more difficult to categorize as specific or non-specific than past events as a result. A rule of thumb adopted while scoring was to reframe future events as past events to see whether reframing as a past event changed its categorization. For instance, a participant’s response “Next year, on New Year’s Eve; my friend and I always go to a party at another friend’s house and like always, we will drink and talk about the previous year” could be reframed as “last year, on New Year’s Eve; my friend and I went to a party like always; we drank and talked about the previous year”. Based on consultation with experts on the AMT, the above example would be categorized as specific.

Other mentions of interrater reliability for the AMT in the literature have been high; for instance, Dalgliesh et al. (2007, p. 27) reported a kappa value of .78, and noted that this was “comparable with previous studies”. Following training, both research assistants independently scored SIT data from five participants. Examination of the correlation between my categorization of responses and theirs based on total number of specific events generated was medium for the first research assistant (r = .54) and large (r = .93) for the second, but proportion of agreement of ratings for individual participants revealed large variability, with agreement as low as .40 and .47 between my ratings and those of each research assistant for the future responses of one participant. As a result of the item-level inconsistency of trial categorization between the research assistants and me, I concluded that neither research assistant had a strong enough understanding of the SIT scoring criteria to proceed with independent scoring. Therefore, I chose to score the SIT alone and revised my plan to examine reliability of the SIT.

To ensure reliability and validity of scoring, Dr. Girard scored data from 15 participants selected from different stages of data collection and scoring, making up approximately 10% of data. Given concern about lower reliability for scoring of future responses, overall reliability as well as reliability for past and future separately were examined using ICC and t-tests of mean differences. The mean ICC using a two-way mixed effects model with absolute agreement (McGraw & Wong, 1996) on the number of specific responses was .96 (95% *CI*: .90-.99), $p < .001$. The absolute mean difference was |0.13| and a t-test of the mean difference was non-significant ($p = .95$). ICC for past responses was .97 (95% CI: .91-.99), $p < .001$, mean absolute difference = |0.13|, $p = .86$. ICC for future responses was .94 (95% *CI*: .84-.98), $p < .001$, mean absolute difference = |0.27|, $p = .84$. Overall, these results highlight that there was high reliability between my identification of specific events and that of Dr. Girard. This suggests that scoring of the SIT was reliable.

Data Analyses

The data were screened for normality and both univariate and multivariate (Mahalanobis' D) outliers; means and ranges and t-test results of clinical and demographic variables for the DG and NDG are summarized in Table 1. Variables with outliers based on standard score (z score > |2.5|) included age, high scores in the DG on clinical measures, and high scores on certain conditions of the MTT tasks. Generally no participant had extreme scores on more than one variable; in 2 cases where participants had high scores on several MTT variables, it was due to high performance across conditions within one of the MTT tasks. To ensure that outliers did not unduly affect the results, the central analyses were conducted with outliers excluded and compared to the results from the analyses with all participants, and the results did not change. Therefore, all data were retained in the analyses.

Prior to conducting the analyses, the SIT data were adjusted to reduce the impact of missing responses. Missed responses were due to participants asking to skip individual items, which they were permitted to do for any portion of the task without disclosing the reason, as recommended by the Ryerson University REB. To reduce the impact of missed responses on any individual condition, the sum of number of specific responses per condition was divided by the total number of responses, and multiplied by five. This adjustment simply converted the total number of specific responses to a proportion score; the proportion score was then multiplied by five to increase comprehensibility of the scores. All analyses relating to the SIT were conducted using these adjusted scores.

Following data screening and preparation, including MTT task scoring and reliability analyses, examination of group characteristics, and refinement of group criteria, bivariate correlations were used to explore the predicted relations among variables based on previous literature, and to compare the magnitude of the relations between performance on each MTT task condition and depression severity scores on the BDI-II. Secondary analyses were likewise conducted to explore correlations among the additional measures listed above with MTT task performance variables. There were 15 correlations between all six conditions within each task and therefore 36 correlations between tasks. Given the high number of correlations in these analyses, I calculated Bonferroni-corrected alpha values to assist with interpretation of the correlation results. Using family-wise alpha of .05 for each set, I used a Bonferroni-corrected alpha of .05/15 = .0033 for correlations within each task and .05/36 = .0014 when correlating both tasks.

To determine whether performance patterns on MTT task conditions differed between the DG and NDG, I conducted an omnibus three-way mixed-factors Analysis of Variance (ANOVA) for each task type (SIT and FIT) on the between-subjects factor of Depression Status (2 levels: DG, NDG) by two repeated factors of Temporal Directions (Past, Future) and Cue Valence (Negative, Neutral, Positive). Main effects and interactions were also examined using Analysis of Covariance (ANCOVA) using covariates identified through correlation. Interaction effects were examined using planned contrasts (t tests) in line with my a priori hypotheses. I also analyzed patterns of performance on the SIT and FIT tasks by qualitatively examining the main effects and interactions from the ANOVAs and ANCOVAs for each task to determine whether the same or different patterns relating to Temporal Direction and Valence are apparent within each task for each group.

Chapter 5: Results

Correlations

Correlations were examined first split by group and then collapsed across both groups. The pattern of correlations was consistent between groups, so I reported correlations collapsed across groups.

FIT. Correlations between responses to conditions of the FIT are summarized in Table 3. As hypothesized, all 15 possible correlations between the 6 conditions of the FIT were significant and positive and survive Bonferroni correction. Correlation coefficient magnitudes ranged from medium (r = .535, p < .001) to large (r = .818, p < .001). Across Past and Future, same-valence conditions are correlated most highly.

SIT. Correlations between conditions of the SIT are summarized in Table 4. Consistent with the hypotheses, Past conditions of the SIT were significantly positively correlated with medium to large effect sizes (rs > .505, ps < .0033). Contrary to hypotheses, 4 of 9 possible correlations between Past and Future conditions of the SIT were significant based on an uncorrected alpha and negative (rs > |-.252|, ps < .034), and the remainder were not significantly correlated (rs < |-.179|, ps > .136). The pattern was such that Past and Future Negative and Neutral conditions negatively correlated with one another, whereas Positive Past and Future conditions were not significantly correlated.

FIT with SIT. Patterns of correlation within the FIT and SIT are summarized in Table 5. Out of 36 possible correlations, 24 were significant and positive and an additional 2 were trending toward significance. The FIT and SIT conditions were mostly significantly positively correlated with each other, although the magnitude was lower (small to medium effect size) than correlations between conditions within each task. The pattern was such that more significant

Table 3: Correlations between conditions of the FIT

	Negative Past	Neutral Past	Positive Past	Negative Future	Neutral Future	Positive Future
Negative Past						
Neutral Past	.596**					
Positive Past	.647**	.593**				
Negative Future	.735**	.541**	.725**			
Neutral Future	.535**	.818**	.616**	.568**		
Positive Future	.646**	.638**	.796**	.704**	.731**	

** Correlation is significant at the Bonferroni corrected alpha level of .0033 (2-tailed).

Table 4: Correlations between mean Adjusted Total number of events reported per conditions of the SIT

	Negative Past	Neutral Past	Positive Past	Negative Future	Neutral Future	Positive Future
Negative Past						
Neutral Past	.733**					
Positive Past	.695**	.738**				
Negative Future	-.277*	-.252*	-.106			
Neutral Future	-.359**	-.275*	-.179	.718**		
Positive Future	-.068	-.050	.052	.505**	.513**	

* Correlation is significant at the .05 level (2-tailed).

** Correlation is significant at the Bonferroni corrected alpha of .0033 (2-tailed).

Table 5: Correlations between mean Adjusted Total number of events reported per conditions of the SIT and Total Number of Events Reported per condition in the FIT.

	SIT Negative Past	SIT Neutral Past	SIT Positive Past	SIT Negative Future	SIT Neutral Future	SIT Positive Future
FIT Negative Past	.093	.134	.121	.335*	.314*	.352*
FIT Neutral Past	.283*	.223	.254*	.272*	.278*	.248*
FIT Positive Past	.167	.252*	.253*	.276*	.275*	.299*
FIT Negative Future	.146	.113	.127	.241*	.252*	.193
FIT Neutral Future	.313*	.258*	.256*	.236*	.259*	.218
FIT Positive Future	.055	.135	.159	.373**	.415**	.319*

Note: SIT = Specificity Instruction Task, FIT = Fluency Instruction Task.
* Correlation is significant at the .05 level (2-tailed).
** Correlation is significant at the Bonferroni corrected alpha of .0014 (2-tailed).

correlations were found between the Future conditions of the SIT and Past and Future conditions of the FIT whereas the Past conditions of the SIT correlated inconsistently with the FIT.

MTT and Other Measures. Exploratory correlational analyses were conducted to examine hypothesized correlations between constructs in the sample based on the literature and to identify potential covariates for use in the main analyses. Due to the high number of correlations computed, I calculated a Bonferroni corrected alpha value due to the probability of experiment-wise error. Using alpha of .05, Bonferroni would set a corrected alpha of: .05/69 = .00072 for correlations from each MTT task between conditions and clinical and cognitive measures. Correlations between the FIT and clinical and cognitive measures can be found in Table 6 and correlations between the SIT and clinical and cognitive measures can be found in Table 7. Contrary to the hypotheses, there was no significant correlation between the FIT conditions and BDI-II. There were significant positive correlations between 5 of 6 conditions of the FIT and a trend toward a significant positive correlation between the other (Negative Past) condition and FAS and Animals. There were significant positive correlations between the BDI-II and other often linked measures, including the BAI, RRS, BHS, and WRAT (see Table 6). The Negative Future condition of the FIT correlated significantly with the BAI and the Neutral Future condition of the FIT correlated significantly with the BHS; the Neutral Past condition of the FIT had a small magnitude correlation with the BHS. Otherwise, no significant correlations were found between conditions of the FIT and clinical or cognitive measures administered.

Table 6: Correlations between clinical measures and FIT and BDI-II

		BDI-II	Negative Past Events	Neutral Past Events	Positive Past Events	Negative Future Events	Neutral Future Events	Positive Future Events
BDI-II	*r*		.059	.031	-.002	.176	.061	.039
	p		.626	.794	.987	.141	.613	.745
	N		71	71	71	71	71	71
Participant Age	*r*	-.031	.211	.126	.158	.245*	.067	.150
	p	.797	.080	.300	.192	.041	.584	.216
	N	70	70	70	70	70	70	70
FAS Raw Score	*r*	.132	.223	.289*	.322**	.442**	.340*	.457**
	p	.278	.065	.016	.007	<.001	.004	<.001
	N	69	69	69	69	69	69	69
Animals Raw Score	*r*	-.035	.204	.370**	.299*	.379**	.272*	.294*
	p	.774	.093	.002	.013	.001	.024	.014
	N	69	69	69	69	69	69	69
Longest Digit Span Forward Score	*r*	.104	.017	.066	.071	.048	.156	.027
	p	.403	.892	.594	.570	.701	.206	.827
	N	67	67	67	67	67	67	67
Digit Span Forward Score	*r*	.162	.072	.124	.116	.113	.201	.096
	p	.195	.564	.322	.353	.365	.106	.445
	N	66	66	66	66	66	66	66
BAI	*r*	.700**	.157	.033	.049	.257*	.058	.084
	p	<.001	.199	.789	.690	.033	.637	.495
	N	69	69	69	69	69	69	69
BHS	*r*	.481**	.007	.216	.043	.169	.263*	.094
	p	<.001	.954	.081	.730	.174	.033	.454
	N	66	66	66	66	66	66	66
RRS	*r*	.598**	.050	-.008	.013	.095	.098	.052
	p	<.001	.694	.952	.920	.452	.438	.680
	N	65	65	65	65	65	65	65
WRAT Reading Test	*r*	.246*	.033	.224	-.012	.152	.146	.083
	p	.047	.792	.071	.926	.224	.242	.506
	N	66	66	66	66	66	66	66

Note: FIT = Fluency Instruction Task; BDI-II = Beck Depression Inventory, Second Edition; BAI = Beck Anxiety Inventory; BHS = Beck Hopelessness Scale; RRR = Ruminative Response Scale; WRAT = Wide Range Achievement Test.
*. Correlation is significant at the .05 level (2-tailed).
** Correlation is significant at the Bonferroni corrected alpha of .00072 (2-tailed)

Similarly, the SIT was not significantly correlated with the BDI-II. In contrast to the FIT, only the number of specific Negative Future events on the SIT correlated with the FAS and there was a trend toward a significant correlation between the LDSF and the number of specific events reported on the Past Positive condition of the SIT (results summarized in Table 7). Number of specific Positive Future events reported correlated negatively with the RRS.

Table 7: Correlations between clinical measures and SIT and BDI-II

		BDI-II	Negative Past Events	Neutral Past Events	Positive Past Events	Negative Future Events	Neutral Future Events	Positive Future Events
BDI-II	*r*		.040	.023	.144	.104	-.016	-.131
	p		.738	.850	.232	.389	.896	.276
	N		71	71	71	71	71	71
Participant Age	*r*	-.031	.017	.049	.007	.134	.119	.040
	p	.797	.889	.684	.954	.268	.327	.740
	N	70	70	70	70	70	70	70
FAS Raw Score	*r*	.132	.076	.025	.062	.262*	.187	.029
	p	.278	.535	.842	.616	.030	.123	.815
	N	69	69	69	69	69	69	69
Animals Raw Score	*r*	-.035	-.025	-.049	-.002	.120	.135	.011
	p	.774	.840	.691	.985	.324	.269	.930
	N	69	69	69	69	69	69	69
Longest Digit Span Forward Score	*r*	.104	.135	.130	.218	.055	-.014	.059
	p	.403	.277	.296	.076	.661	.910	.634
	N	67	67	67	67	67	67	67
Digit Span Forward Score	*r*	.162	.138	.063	.130	.158	.086	.039
	p	.195	.268	.613	.297	.205	.491	.757
	N	66	66	66	66	66	66	66
BAI	*r*	.700**	.058	.037	.129	.096	-.006	-.097
	p	<.001	.636	.761	.291	.431	.960	.429
	N	69	69	69	69	69	69	69
BHS	*r*	.481**	.117	.047	.137	.017	.024	.061
	p	<.001	.351	.705	.273	.891	.850	.626
	N	66	66	66	66	66	66	66
RRS	*r*	.598**	-.004	.144	.053	.055	-.139	-.256*
	p	<.001	.972	.251	.674	.665	.269	.040
	N	65	65	65	65	65	65	65
WRAT Reading Test	*r*	.246*	.191	.011	.173	.164	.161	-.085
	p	.047	.124	.927	.165	.187	.196	.496
	N	66	66	66	66	66	66	66

Note: SIT = Specificity Instruction Task; BDI-II = Beck Depression Inventory, Second Edition; BAI = Beck Anxiety Inventory; BHS = Beck Hopelessness Scale; RRR = Ruminative Response Scale; WRAT = Wide Range Achievement Test.

*. Correlation is significant at the .05 level (2-tailed).
** Correlation is significant at the Bonferroni corrected alpha of .00072 (2-tailed).

Main Analyses

FIT. Group means per condition of the FIT can be found in Table 8, and the results have been plotted in Figure 4. An ANOVA was conducted to examine the effects of Valence (Negative, Neutral, Positive), Temporal Direction (Past, Future), and Group (NDG, DG) on the number of events reported on the FIT. There was a main effect of Valence, $F(2, 138) = 88.27$, $p < .001$, partial eta squared, $\eta^2_p = .561$, and there was a trend toward an interaction between Valence and Temporal Direction, $F(2, 138) = 2.38$, $p = .098$, $\eta^2_p = .033$. There was no significant main effect of Group, $F(1,69) = 0.576$, $p = .450$, $\eta^2_p = .008$, or Temporal Direction, $F(1, 69) = 2.116$, $p = .150$, $\eta^2_p = .030$. There were no significant interactions between Group and Valence, $F(2, 138) = 1.21$, $p = .300$, $\eta^2_p = .017$, or Group and Temporal Direction, $F(1, 69) = 0.002$, $p = .962$, $\eta^2_p < .001$. Lastly, there was no significant three-way interaction between Group, Temporal Direction, and Valence, $F(2, 138) = 0.50$, $p = .951$, $\eta^2_p < .001$.

Paired samples t-tests confirmed that there were significant differences between each Valence condition. Participants reported significantly more events in the Neutral than Negative condition, $t(70) = 3.85$, $p <.001$, more events in the Positive than Negative condition, $t(70) = 10.34$, $p <.001$, and more events in the Positive than Neutral condition, $t(70) = 12.61$, $p <.001$.

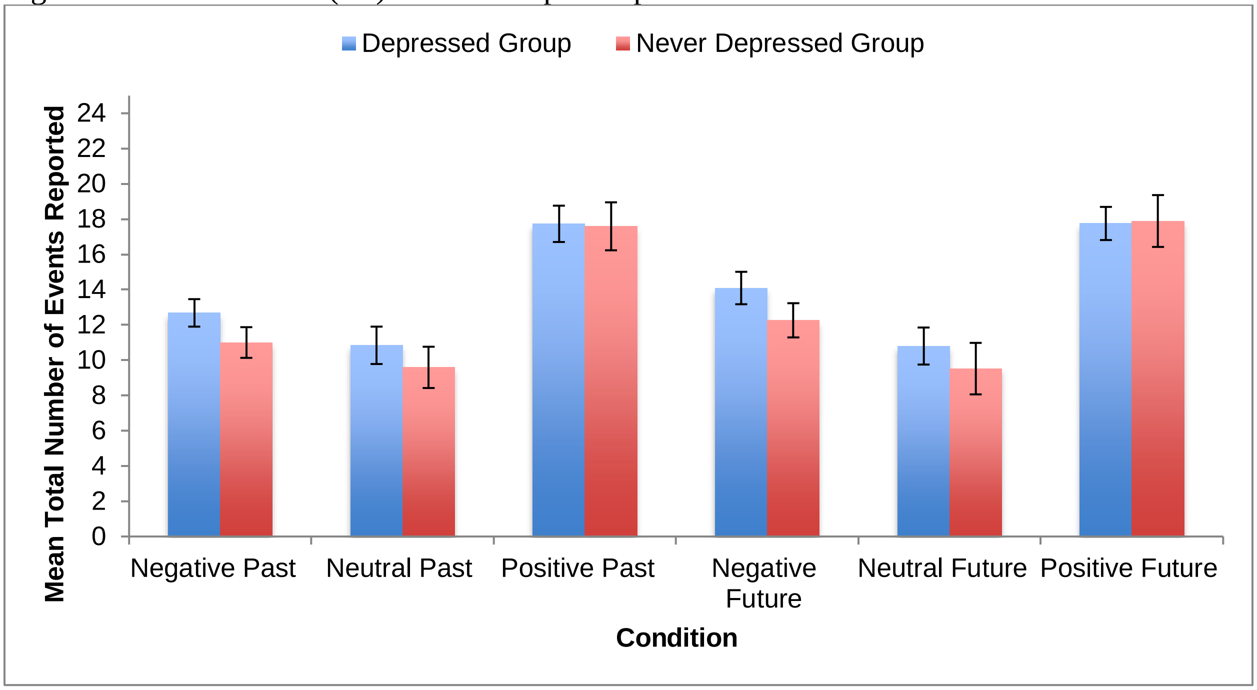
Because most fluency instruction tasks in the literature do not include a Neutral valence condition, an ANOVA with only Positive and Negative Valence conditions was conducted for closer comparison. As in the ANOVA with three Valence conditions, there was a main effect of Valence, $F(1,69) = 112.812$, $p < .001$, $\eta^2_p = .620$. In addition, the main effect of Temporal Direction reached significance, $F(1,69) = 4.485$, $p = .038$, $\eta^2_p = .061$, and there was a trend toward a Valence by Group interaction, $F(1,69) = 3.188$, $p = .079$, $\eta^2_p = .044$. No other main effect or interactions were significant. Despite the trending interaction between Valence and Group a follow-up comparison of the total number of events reported across both Negative valence conditions, where examination of the means suggests that the difference was largest, yielded a small-medium effect size but failed to reach significance, $t(69) = -1.450$, $p=.152$, $d = 0.359$.

Table 8: FIT mean total number of events reported per condition per Group

	Group	N	Mean	Std. Deviation
Negative Past Events	NDG	27	11.00	4.50
	DG	44	12.68	5.20
Neutral Past Events	NDG	27	9.59	6.07
	DG	44	10.84	7.06
Positive Past Events	NDG	27	17.59	7.06
	DG	44	17.73	6.86
Negative Future Events	NDG	27	12.26	5.06
	DG	44	14.09	6.08
Neutral Future Events	NDG	27	9.52	7.56
	DG	44	10.80	7.00
Positive Future Events	NDG	27	17.89	7.65
	DG	44	17.75	6.21
Total Past Events	NDG	27	38.19	15.24
	DG	44	41.25	16.49
Total Future Events	NDG	27	39.67	18.49
	DG	44	42.64	16.79
Total Negative Events	NDG	27	23.26	9.12
	DG	44	26.77	10.36
Total Neutral Events	NDG	27	19.11	12.94
	DG	44	21.64	13.47
Total Positive Events	NDG	27	35.48	13.77
	DG	44	35.48	12.52
Total Events	NDG	27	77.85	33.01
	DG	44	83.89	32.23

Note: NDG = Never Depressed Group; DG = Depressed Group.

Figure 4: Mean numbers (SE) of events reported per condition of the FIT



SIT. Group means per condition of the SIT, adjusted to account for missing values, can be found in Table 9, and the results have been plotted in Figure 5. An ANOVA was conducted to examine the effects of Valence (Negative, Neutral, Positive), Temporal Direction (Past, Future) and Group (NDG, DG) on the number of specific events reported on the SIT. There was a main effect of Valence, $F(2, 138) = 11.707$, $p < .001$, $\eta^2_p = .145$. There were no significant main effects of Temporal Direction, $F(1, 69) = 0.082$, $p = .775$, $\eta^2_p = .001$, or Group, $F(1, 69) = 0.018$, $p = .894$, $\eta^2_p > .001$. There was no significant Temporal Direction by Group interaction $F(1, 69) = 0.404$, $p = .572$, $\eta^2_p = .006$, Valence by Group interaction, $F(2, 138) = 0.208$, $p = .812$, $\eta^2_p = .003$, or Group by Valence by Temporal Direction interaction, $F(2, 138) = 0.431$, $p = .651$, $\eta^2_p = .006$. Exclusion of the Neutral condition (as done for the FIT) had no effect on the results: there was only a main effect of Valence on the SIT, $F(1 69) = 17.574$, $p < .001$, $\eta^2_p = .203$.

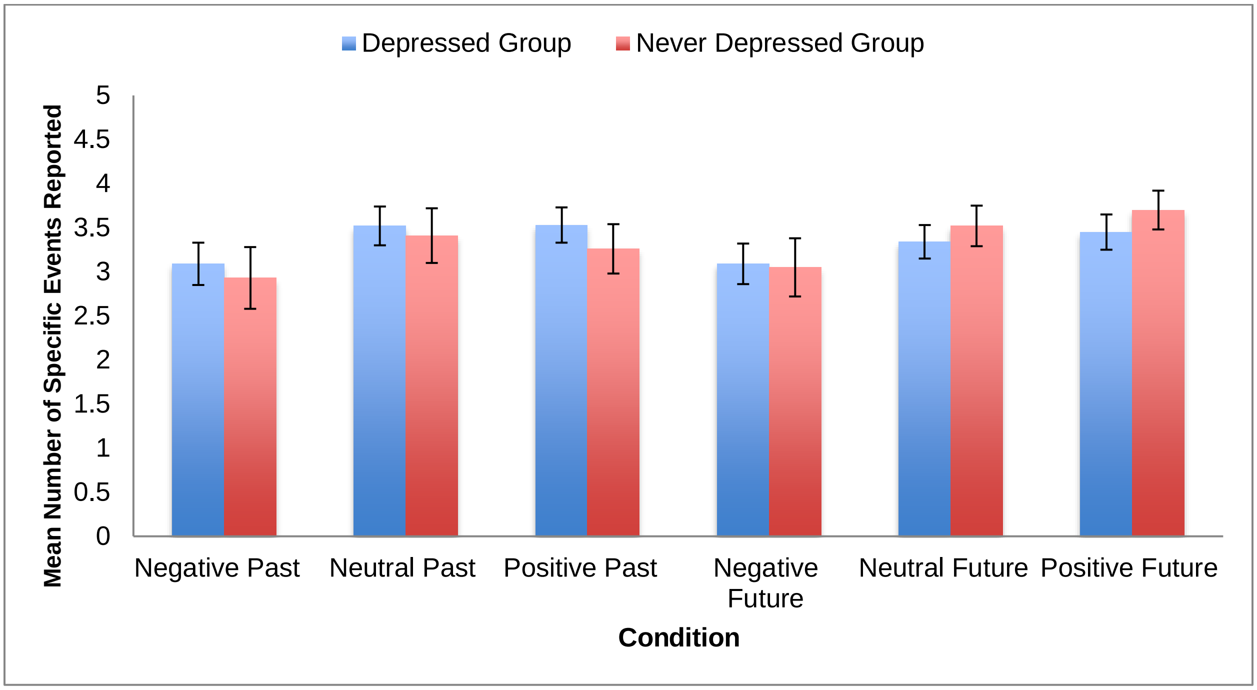
Paired samples t-tests revealed that participants reported significantly more Positive specific events than Negative specific events, $t(70) = -4.237$, $p < .001$, and more Neutral specific events than Negative specific events, $t(70) = -3.998$, $p < .001$. The difference between the number of Positive and Neutral specific events reported was not significant, $t(70) = -0.459$, $p = .648$.

Table 9: SIT adjusted mean number of specific events reported per condition per Group

	Group	N	Mean	Std. Deviation
Specific Past Negative Events	NDG	27	2.93	1.80
	DG	44	3.09	1.58
Specific Past Neutral Events	NDG	27	3.41	1.60
	DG	44	3.52	1.44
Specific Past Positive Events	NDG	27	3.26	1.46
	DG	44	3.53	1.34
Specific Future Negative Events	NDG	27	3.05	1.73
	DG	44	3.09	1.53
Specific Future Neutral Events	NDG	27	3.52	1.22
	DG	44	3.34	1.27
Specific Future Positive Events	NDG	27	3.70	1.14
	DG	44	3.45	1.35
Specific Past Events	NDG	27	9.59	4.37
	DG	44	10.14	3.94
Specific Future Events	NDG	27	10.27	3.43
	DG	44	9.88	3.58
Specific Negative Events	NDG	27	5.97	1.95
	DG	44	6.17	1.98
Specific Neutral Events	NDG	27	6.93	1.59
	DG	44	6.86	1.72
Specific Positive Events	NDG	27	6.96	1.99
	DG	44	6.98	1.91
Total Number of Specific Events	NDG	27	19.86	4.56
	DG	44	20.02	4.89

Note: NDG = Never Depressed Group, DG = Depressed Group.

Figure 5: Mean numbers (SE) of specific events reported per condition of the SIT



Covariate Analyses. Guided by the exploratory correlational analyses, several covariates that were correlated with MTT task performance or BDI-II score were included alone or in combination in ANCOVAs of the FIT and SIT, including the FAS, BAI, BHS, Animals, and WRAT Reading score. Since FAS and Animals were significantly correlated (r = .363, p = .001) but the magnitude of correlation between FAS and MTT measures was marginally higher, FAS was selected for inclusion as a covariate in an ANCOVA for the FIT. For the FIT, inclusion of FAS as a covariate revealed a main effect of Valence, $F(2, 132) = 3.25$, $p = .042$, $\eta^2_p = .047$, and a significant main effect of Temporal Direction, $F(1,66) = 6.69$, $p = .012$, $\eta^2_p = .092$. For the SIT, inclusion of covariates did not reveal any significantly different main effects or interactions. None of the ANCOVAs run revealed significantly different results pertaining to Group for either the FIT or SIT.

Group Definition

Early analyses revealed that no group differences between the DG and NDG were evident. To rule out the possibility that group membership criteria may have been confounding or masking predicted main effects of group, I conducted exploratory analyses to determine possible confounds within each group and explored the study hypotheses using several alternative group classification criteria schemes, including using BDI-II score of 20 or above (indicative of moderate depression severity based on the manual) as criteria for the DG, using either BDI-II score or MINI diagnosis solely to determine group membership, and eliminating stringent criteria for the NDG including self-reported history of learning disability, recent recreational drug use or neurological condition. None of the alternative schemes resulted in main or interaction effects of group on any of the central study tasks. Therefore, the group criteria outlined in the Participants section were retained.

Chapter 6: Discussion

Review of Hypotheses

Objectives of this study included replicating findings from past literature related to future thinking and AM in depression, replicating findings that the MTT Model extends to fluency and specificity instruction tasks in a depressed sample, and the CaR-FA-X Model, and to determine whether temporal direction accounted for inconsistent valence patterns in AM and future thinking in depression. In summary, I hypothesized that measures of depression, hopelessness and ruminative tendency would correlate with MTT task performance. Based on the FT literature and CaR-FA-X Model, I hypothesized that there would be a significant main effect of Group for the SIT and an interaction between Group and Valence on the FIT. Based on the MTT Model, I hypothesized that the FIT and SIT would demonstrate significant positive correlations between all conditions within each task and all conditions between tasks, with larger magnitude correlations within than between each task, and within temporal direction and valence conditions than between.

Consistent with the hypotheses relating to the MTT Model, all conditions within the FIT were significant and positively correlated with medium to large correlation magnitudes. There were significant positive correlations of medium to large magnitude within each temporal direction condition of the SIT, but across temporal direction conditions, there were significant negative correlations of small magnitude between the Neutral and Negative Past and Future conditions, and no significant correlation between Positive Past and Future conditions. In addition, whereas most conditions of the FIT and SIT were significantly positively correlated with effect sizes ranging from small to medium, the Negative Past and Future and the Positive Future conditions of the FIT were not significantly correlated with Past conditions of the SIT.

Overall, most correlations predicted by the MTT Model were found in this sample and importantly some were not; implications of this finding will be discussed later in this section.

Contrary to the hypotheses, the BDI-II, the primary measure of depression, did not correlate with performance on any conditions of the MTT tasks, although it was moderately to strongly correlated with other clinical measures. There were some small magnitude correlations between clinical measures such as the BAI, BHS and RRS and individual conditions of the FIT or SIT.

In line with the hypotheses, there was a main effect of Valence and a marginal interaction between Valence and Temporal Direction in the FIT. However, this study did not replicate the finding of a Group by Valence interaction effect; examination of means indicates that any Group effects were in the direction opposite of what would be predicted by the literature, in that the DG reported more details than the NDG, rather than less. For the SIT, there was a main effect of Valence; unlike what was predicted by the CaR-FA-X Model, there was no significant main effect of Group. Examination of means suggests that although the direction of any possible effect of Group could be consistent with the pattern predicted by the CaR-FA-X Model (fewer events reported by the DG than the NDG on some conditions), based on effect size, the clinical utility of this effect would be limited. In sum, the results of this study raises questions about findings from previous studies of reduced positive future fluency and Overgeneral AM in depression.

Valence and Temporal Direction

As summarized in the literature review, due to differences in task instructions and paucity of direct comparison between these tasks in clinical samples, it was unclear whether differing Valence effects reported in previous studies were due to differing task sensitivity or whether the tasks assess distinct cognitive processes. Comparison of within-task patterns of Valence and

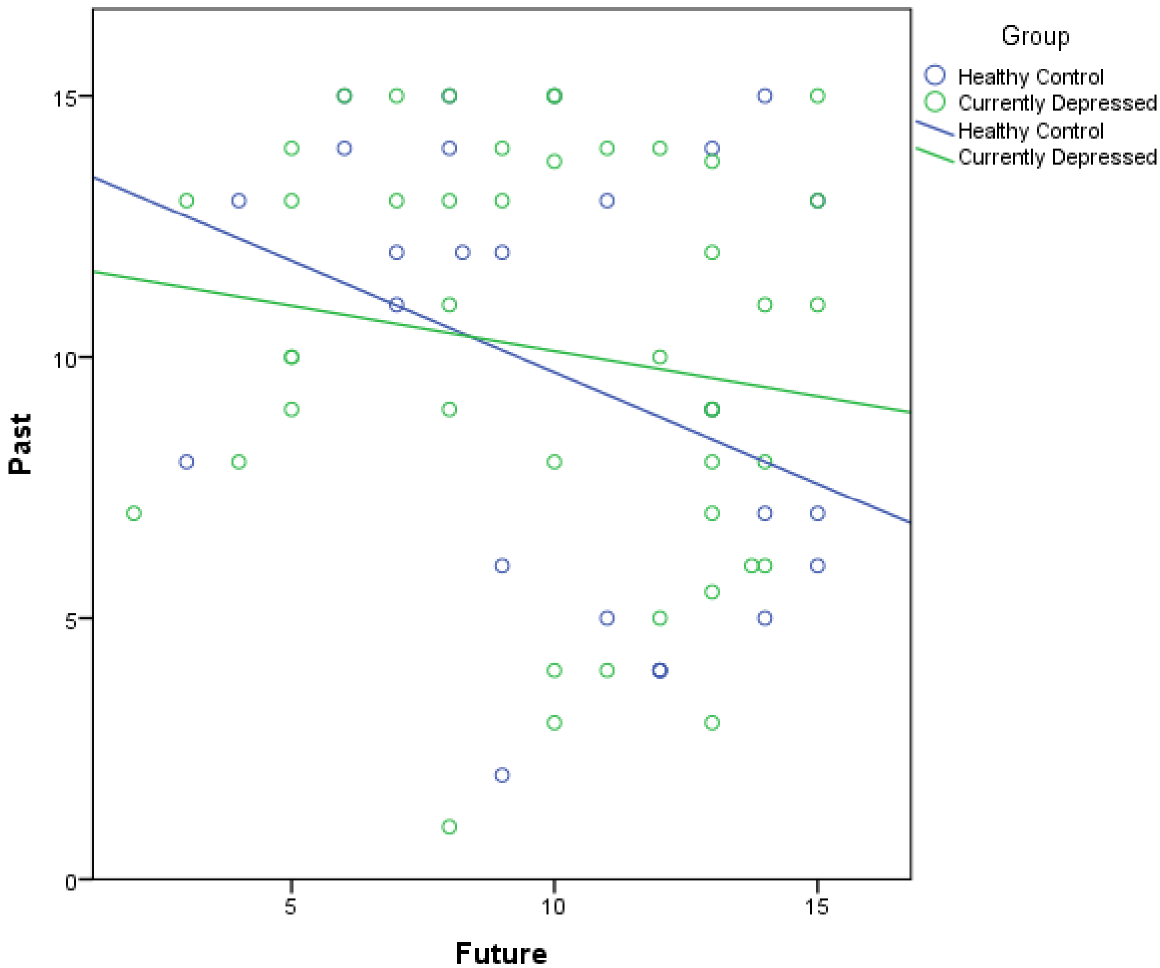
Temporal Direction conditions between the tasks (Figures 4 and 5) shows a facilitation effect of emotion on the FIT such that more events were reported in the Positive and Negative than Neutral conditions across Past and Future. The SIT showed a pattern related to Valence such that more events were reported in the Neutral and Positive than Negative condition, and more events were reported in the Positive than Neutral condition. Consistent with the MTT literature, in both tasks, more Positive than Negative events were reported across both Past and Future, indicating a possible avoidance of negative thoughts or a positivity bias in FIT as well as SIT. Importantly, while the pattern of results *between* the tasks was not consistent, the pattern of results was similar across both temporal direction conditions *within* each task, suggesting that Temporal Direction has not been confounding previous research on MTT in clinical samples. However, given that there were no Group differences found in my dissertation, this hypothesis needs to be re-tested in a study in which Group differences are observed. If Temporal Direction can be ruled out as the cause of differing Valence effects between the FIT and SIT, it can be concluded that task instructions account for the difference in prior research.

There was a small negative correlation between Past and Future conditions of the SIT. Given the unexpected direction of this correlation, I chose to explore this finding further by examining correlations within the SIT by group (see scatterplot in Figure 6) to understand whether this finding was related to depression in this sample. Examination of the negative correlations between Past and Future conditions of the SIT by Group confirms small magnitude negative correlations in both groups ($r = -.33$ for the NDG, $r = -.16$ for the DG), suggesting that this pattern is caused by an individual difference factor or strategy unrelated to depression. None of the other characteristics measured in this study, including anxiety, ruminative tendency, number of past traumas, or executive functioning, correlated uniquely to Past or Future conditions, so it is unlikely that any of these characteristics account for this finding. This suggests that there are non-clinical individual differences that affect performance on MTT Tasks.

An individual difference variable that has not been considered in the CaR-FA-X or MTT Models is trait-like preference, bias, or tendency regarding Time Perspective, described by Zimbardo and Boyd in 1999 as “subjective conception of focusing on various temporal categories or time frames when making decisions and taking action”. Zimbardo and Boyd developed the Zimbardo Time Perspective Inventory (ZTPI; 1999, p. 129) to assess this trait. Based on a factor analysis, Time Perspective has five dimensions: Past-Negative, Past-Positive, Present-Hedonistic, Present-Fatalistic and Future factors. An individual may adopt any of these perspectives regarding time simultaneously. The Past-Negative factor includes having a negative view of one’s past; for instance, items on this subscale include: “I think about the bad things that have happened to me in the past” and “I think about the good things that I have missed out on in my life”. The Past-Positive factor captures a perspective that includes warm, nostalgic feelings about one’s past; examples of items on this subscale include: “I feel nostalgic about my childhood” and “I like family rituals and traditions that are regularly repeated”. The Present-Hedonistic factor encompasses what the authors describe as a “devil may care” attitude about life, including a focus on present enjoyment over long-term gain. Examples of items assessing the Present-Hedonistic factor include “Taking risks keeps my life from becoming boring”, and “I do things impulsively”. The Present-Fatalistic subscale relates to feelings of hopelessness and helplessness to affect outcomes of events in one’s life. Examples of items on the Present-Fatalistic subscale include: “Often luck pays off better than hard work” and “My life path is controlled by forces that I cannot influence”. Lastly, the Future factor measures tendency to be conscientious and motivated by goals. Examples of items on this subscale include: “It upsets me to be late for appointments", and "I complete projects on time by making steady progress". Response patterns on the ZPTI, particularly high endorsement of items on the Present-Hedonistic and Present-Fatalistic subscales, and low endorsement of items on the Future subscale have been linked to substance abuse (Keogh, Zimbardo & Boyd, 1999), coping with homelessness, and achievement in academic and occupational settings. Future studies should explore the relevance of Time Perspective to the MTT and CaR-FA-X Models by administering the ZPTI alongside the FIT and SIT.

Other limitations in interpreting negative correlations between Past and Future conditions of the SIT stem from lack of specificity in the MTT Model. The MTT Model is based on the premise that FT and AM rely on the same underlying neurological and cognitive processes. However, AM and FT may have different functional and emotional significance in healthy people. It is equally possible that the MTT Model holds true on a neurological level, but that additional processes, such as biases in Time Perspective, mediate performance on behavioural tasks, such as the SIT. Overall, the lack of positive correlation between Past and Future conditions of the SIT and between SIT and FIT suggest that the FIT and SIT have different task demands and are likely reliant on differing cognitive processes.

Figure 6: Scatterplot of total number of specific past and future events reported per group on the SIT



Interestingly, as depicted in Figure 5, although ANOVAs and t-tests show that there was no significant difference between the groups on the SIT, there was a pattern such that the DG reported more events than did the NDG in the Past conditions, but fewer events than the NDG in two of three Future conditions. This pattern is in contrast with the finding that the DG performed better than the NDG in all other MTT task conditions. Both the pattern associated with past and future specificity in depression and seemingly better performance in most conditions by depressed participants warrants closer examination in future studies.

Lastly, although the predicted relations between MTT and depression were not found, several published studies have not reported group effects (see Sumner et al., 2010), and other predicted correlations within and across MTT tasks were detected, indicating that the methodology used at least partially replicated other literature. In fact, my dissertation replicated and expanded on findings from Sarkohi et al. (2011) that there is no correlation between the Negative Valence condition of the AMT, the specificity instruction task used in that study, and Future conditions of the Future Thinking Task (FTT), the fluency instruction task used, or between depression and either of these tasks in this study. My dissertation replicated the finding that performance on the Past Negative condition of the SIT did not correlate to the Future Negative condition of the FIT and extends this finding to the Negative Past condition of the FIT, supporting the conclusion that these tasks are somewhat but not entirely overlapping, and that perhaps a unique process mediates performance on the Negative Valence condition of specificity instruction tasks in particular. In addition, my dissertation partially replicated the finding of a positive correlation between the BAI and the Negative Future condition of the FIT (MacLeod & Byrne, 1996), although the correlation did not survive Bonferroni correction. Thus, although the relation between MTT and depression predicted in my hypotheses was not replicated, my dissertation replicates other findings from the literature.

Replication Issues

The lack of replication of differences in MTT processes in depression in the current study suggests at least two possible conclusions: first, that some part of the methodology used in this study was different from that of previous studies in a way that masked a true relationship between MTT processes and depression; second, that there is a systematic error, omission, or bias in the previous literature. If true, either of these conclusions would be surprising given the amount of effort I expended to ensure that the tasks were administered in the same manner in which they have been administered in other studies. However, in view of the evidence, I think that these conclusions are both likely true to some extent. Recently, meta-research about reproducibility of psychological research reported that 61 of 100 attempts at replication, even with high levels of collaboration with the original authors, failed, highlighting both the low rate of replication in psychology studies as well as the importance of conducting replication studies and publishing non-significant findings (Open Science Collaboration, 2015). As noted in this article, failed replications are an essential part of research as they can clarify true effect sizes, identify type I errors, identify issues with transparency or clarity of study protocol, or indicate unknown moderators of the effect. In the current study, I believe that any of the above factors could account for the results.

To support this position, I will summarize evidence that the current pattern of findings is not easily attributable to a simple methodological error unique to this study, review and evaluate alternative possible explanations, including limitations and weaknesses of both the current study and literature, and make recommendations for future research. Potential methodological issues discussed in the paragraphs to follow include lack of methodological clarity and consistency regarding scoring and prompting practices in the literature, unknown moderators of MTT differences in depressed samples (cultural factors, total length of depression, age of onset, number of depressive episodes), and conceptual issues associated with application of the MTT Model to the FIT and SIT.

The first important point to consider is whether the lack of replication of previous findings is due to methodological error. Based on my experience conducting this study, I believe that not error, but lack of methodological transparency in the literature may have contributed to the lack of replication. As discussed in the methodology section, I made every effort to understand and mimic methodology from previous studies with the goal of replicating the findings related to future fluency and AM specificity in depression. Although the methodology described in published papers seems transparent, in administering and scoring the tasks, several ambiguities quickly became apparent. Decisions that I struggled with in the design that could be linked to lack of replication of reduced memory specificity in the SIT included where to recruit my sample, which methodology to use to detect and characterize depression, whether to exclude participants with comorbid conditions such as anxiety disorders or PTSD, how many and which cues to use, how long to allow for each trial, when it was appropriate to prompt participants and how to use prompting in scoring, how to treat omissions and repeats, and how to reliably distinguish ‘specific’ from ‘overgeneral’ in scoring. Ambiguities relating to the FIT were fewer, but how to count distinct events and repeats was unclear based on the literature.

Many of the issues with the SIT only became apparent when administering the task; based on description in the literature and consultation with experts in the field, there is variability in administration procedures, and methodology has changed over time. For instance, papers reviewing reduced memory specificity in depression have noted that although the literature as a whole supports this finding, individual studies have yielded inconsistent results. Given variability in published methodology, some of which has arisen directly from failed attempts to replicate this finding (e.g. Debeer, Hermans, & Raes, 2009), it is also possible that this literature suffers from publication bias. One potential issue highlighted in reviews of this literature and my dissertation is lack of clarity about the population affected by MTT differences.

Sampling issues. First, it is clear that not all people with depression have reliably demonstrated Overgeneral AM in previous studies. Examination of results from meta-analyses

(Sumner et al., 2010; van Vreeswijk & de Wilde, 2004) confirms this inconsistency, but what is distinct about samples used in studies detecting this effect is unclear. In my study, although I recruited from an undergraduate sample and the BDI-II scores were positively skewed, the mean BDI-II score was 22 (moderate), 18 of 44 participants endorsed some degree of *current* suicidality, as measured by endorsement of an item assessing suicidality on the BDI-II, and 31 of 44 endorsed significant hopelessness on the BHS, as defined by a score greater than or equal to 3. If severity is an important predictor, I would anticipate that a sample with these characteristics, if depression were not severe enough to detect group differences, would at least have sufficient power to detect some magnitude of correlation between MTT task conditions and depression, but this was not the case. In recognition of this possibility, I conducted exploratory analyses using more stringent criteria to define depression, but the results of the analyses did not change. Further, several other studies in this literature have reported significant effects with samples with similar BDI-II scores, including the original study describing MEST (Raes, Williams, &Hermans, 2009).

It could be argued that findings from the literature were not replicated because the DG in this study was somehow atypical, but overall the data do not support this hypothesis. Differences between my sample and those of previous studies that detected the effects I was seeking to replicate, in addition to depression severity and suicidality, might include recruitment source (undergraduates vs. inpatient), age, or cultural factors. In a meta-analysis of the AMT, which has been more thoroughly studied than the FTT, age was not found to be a moderator of memory specificity, and self-reported depressed mood was a significant moderator of specificity (van Vreeswijk & de Wilde, 2004). In the same study, the authors noted that the clinical samples used reported depression scores two to three standard deviations higher than the non-clinical control samples used. In my dissertation, the difference in depression scores between the groups met this criterion (see Table 1). Given the characteristics of my sample, it therefore does not seem likely that differences in age or depression severity account for the findings. If recruitment source is a moderator in this case, it would be unclear as to why, if the sample has depression of similar severity to that of a sample recruited from a clinical setting based on existing models of MTT in depression. One difference that could arise based on recruitment source is motivation for participation. Participants in both groups of this study were offered course credit for participation, whereas participants recruited from clinical settings are usually offered money. A related difference could include intrinsic interest in research.

Further supporting the generalizability, I found significant correlations between the BDI-II and other self-report clinical measures such as the BAI, BHS, RRS and THS. The magnitude of correlation between these measures was moderate to high, and in the expected direction. Further indications that this sample was generalizable include that the DG and NDG differed significantly on the above clinical measures in the predicted directions, performed in the average range on the neuropsychological measures administered, and that based on the reliable digit span, which was administered late in the study session, participants were still exerting adequate effort.

Whether cultural factors could have affected the results is difficult to determine because previous studies have not examined possible cultural differences or reported the cultural makeup of their samples. Although I did not collect information about the cultural identity of participants in this study, participants were asked whether they were fluent in English, meaning that English is their primary language or was learned before age 5; in my sample, 20 reported that it was not their primary language. Based on the phrasing of this question, it is probable that this is an underestimate of participants coming from diverse cultural backgrounds because, for example, second generation immigrants would have learned English at school before age 5 even if a different language was spoken in the home, and this question does not capture participants from different cultural backgrounds who speak English as a primary language. Although a high proportion of the sample identify English as a second language, WRAT Reading scores and the recruitment source affirm that English fluency is not likely to have affected the results.

Overall, there is no clear indication in the results to suggest that this sample is somehow unique. One additional possible explanation is that failure to replicate group differences is attributable to sampling error in either the DG or NDG. Future research is needed to test whether cultural factors or factors related to depression severity are moderators of MTT changes in depression.

Prompting. The second issue apparent from the literature and firsthand experience administering these tasks is inconsistent use of prompting. As noted in AMT instructions obtained through personal correspondence from two leading researchers in memory specificity in Appendix C, researchers are instructed to prompt if it is unclear whether a response is specific. However, there is no clear operational definition of clarity, resulting in prompting based on use of personal judgment influenced by factors such as speed with which the participant responds, and phrasing.

In this study, I prompted participants on a trial-by-trial basis immediately following the response if the response was not clear. The benefits of this strategy include eliminating the possibility that the participant gains extra time during which to elaborate upon or generate a specific event throughout the remaining trials, thereby theoretically increasing the reliability of use of time to respond as an indicator of specificity of an event, and the ability to provide corrective feedback if the participant does not understand the task instructions. Undesirable consequences of this method include potentially differing levels of feedback and task duration between participants.

In choosing how to administer the task, the consideration that resulted in the decision to use liberal prompting was the perception that the construct of overgeneral memory as laid out in the SMS Model relates not to clarity of reporting, motivation or strategy, but reflects an underlying cognitive process related to memory systems; if so, prompting ought not to change the nature of participants' responses, and should provide more information to researchers upon which to base the judgment of whether an event reported was specific or not. Further, although it was not feasible to be entirely blind to the presumed participant group membership based on prescreening data due to the recruitment procedure through Sona, the data collection schedule and pace was hectic and the MTT tasks were completed before clinical measures, reducing the likelihood that participants in each group received different prompting strategies based on experimenter bias.

If prompting in this study did account for nullification of the Overgeneral AM effect, it would suggest that instead of a change in retrieval or generation of events in depression, the effect is due to a shallow reporting bias. This would also be inconsistent with the CaR-FA-X model, which proposes curtailing of the retrieval process, rather than reporting bias, as the mechanism underlying the effect. At minimum, this study would suggest that there is no impairment in recall of specific events in depression *per se*, but this study is not able to rule out the possibility of a reporting bias in depression.

Another consideration is that if simply increasing prompting negates the effect, it is questionable whether any impairment or bias was ever very enduring or significant. If this were the case, it would be expected that only studies using samples with severely depressed people would detect the Overgeneral AM effect. This study also highlights the importance of studying potential long-term impact of prompting on event reporting strategies in depression. If the Overgeneral AM phenomenon is due to a bias in reporting rather than an impairment in retrieval, if this bias is indeed central to depression or contributes to development or maintenance of depression, this could suggest that training in meta-cognition about memory retrieval and cognitive avoidance in depression and/or the therapeutic strategy of cognitive restructuring may be as important as training and practice in retrieving specific events, since awareness or preference regarding retrieval strategies could undermine the effectiveness of memory specificity training. Alternatively, if prompting negates Overgeneral AM, this may suggest that Overgeneral AM is a correlate of depression, but not causally implicated in depression.

Task design. Another alternative explanation for the lack of replication in this study is lack of external scoring validity. Although adequate consistency of scoring was attained, as evidenced by reliability analyses, whether the scoring strategies used are in line with what is used consistently in past research is unclear because many quirks regarding scoring of the tasks are not noted in publicly available task documentation. As outlined in the Method section, dilemmas in the scoring of the FIT included distinguishing distinct events within trials and identifying repeats across trials and conditions. Dilemmas in the scoring of the SIT included distinguishing between overgeneral and specific memories, how to manage omissions and repeats in analyses, and the difficulty of operationalizing specific future events. To assist with scoring of the FIT, it was helpful to adopt the attitude of giving the participant the benefit of the doubt concerning possible repeats, and to not over interpret results based on participants' phrasing. Unlike the verbal fluency task upon which they are based, in contrast to what I expected based on the literature, many participants do not list events in point form, but report events using complex sentences and narratives. Based on feedback from MacLeod via personal communication (2015), sentences containing many possible events were scored as a single event to reduce possible researcher inference, but whether this has been done consistently in past studies is unclear. For the SIT, I also adopted the strategy of giving participants the benefit of the doubt if an event was unclear. When participants gave a first response that was unclear and a prompt was given, I used the recommendation that if they responded immediately, it was likely that the participant had been thinking of a specific event during the first response. The most difficult events to score included mundane events, where it was likely that the person would have experienced semantically similar events numerous times, and future events, which were often based on past experience and therefore were also semantically similar to other events.

Another challenge with this study was the use of both past and future conditions in the tasks. Although this has been done with both fluency and specificity instruction tasks in other studies (e.g. Addis et al., 2016; D'Argembeau & Van der Linden, 2004; MacLeod et al., 1997), conceptually, it presented challenges. To illustrate, it is necessary to consider the purpose of each task and definition of the abilities sought to be measured by each. The FIT and similar fluency instruction tasks are designed to measure fluency of recall and/or reporting of autobiographical events and was developed based on phonological and semantic fluency tasks (FAS and Animals). In these tasks, performance is defined by generation of unique words, representing unique ideas. For example, in the FAS, if someone said "shade", then "shaded" and "shading", they would only receive a point for a single answer. Similarly, on Animals, "cat" and "kitten" would not both be scored correct because of their semantic overlap. But because fluency and specificity instruction tasks examine distinct events, and distinct events are differentiated by time rather than overlap of content, it is not possible to reliably distinguish events with absolute certainty. A good example of this would be events such as “walking the dog” or “going to the doctor’s for a checkup”. Because this is a naturally recurring activity, the line between specific and overgeneral is very blurry and adding the factor of temporal direction only makes this distinction more difficult. Someone can recall distinct events related to walking the dog across past and future, but these events may be highly semantically overlapping and their uniqueness is uncertain, particularly when judged by a non-expert or someone besides the person reporting the event, who, paradoxically, is also a non-expert, because identification of distinct events requires a thorough understanding of this construct.

The possibility that greater group differences would have emerged if only data from the first task completed by the depressed sample were examined due to practice or learning effect cannot be dismissed, but given how well the DG performed relative to the NDG, it is unlikely that any differences would have been in the directions predicted by the literature.

Ultimately, the difficulty with both fluency and specificity instruction tasks is that much of the scoring relies on researcher judgment, hand scoring and recording, and the phenomena being measured are subjective. Although efforts were made to standardize administration as much as possible in this study through use of scoring guides, time limits, standard scripts, careful selection of psychometrically equated stimuli, and counterbalancing conditions and presentation of stimuli, as summarized in the previous paragraphs, there remain many confounds in these tasks.

Limitations

Pursuant to the methodological issues surrounding prompting noted in the discussion section, it would have been helpful to be able to characterize use of prompts in this study quantitatively, and I do not think that this has been done in the literature to date. Among data entered for this study, I included several variables about prompting. Data collected included latency to first response, number of prompts given per trial, and number of prompts to follow task instructions given per trial. I recorded latency of response times, but I did not record latency of response times following prompting because there was no apparent need for that during the data scoring and entry phase of the study. I did not analyze these data due to time limitations and because this was not a central research question, but future research could examine the effect of different prompting methodology on outcomes.

Additional variables relating to the MTT tasks that could not be included or analyzed in this study due to testing duration, scoring burden or participant fatigue include subjective valence of events reported, self-relevance of stimuli, temporal distance of events reported, vividness, reporting strategy (liberal or conservative), and reaction time.

Although the depressed sample that I recruited from the Sona undergraduate participant pool reported depression symptom severity consistent with that of a clinical sample, and many met diagnostic criteria for recurrent depression, had hopelessness and/or suicidality, it is possible that samples recruited from clinical settings are qualitatively different from undergraduate samples. Differences between my sample and clinical samples could include duration of depression, number of major depressive episodes, age of onset, socioeconomic status, differences in quality or severity of symptoms not measured by self-report questionnaires, degree of functional impairment, amount of subjective distress, psychological mindedness, social support, cognitive functioning, health status, or a host of other factors. Recruitment setting has never been identified as a moderator of MTT effects in depression and reduced memory specificity has been reported in a number of samples in the past, including subclinically depressed (Dalgleish et al., 2007), dysphoric (Dickson & Bates, 2006), at risk for depression (Young, Bellgowan, Bodurka, & Drevets, 2013) and remitted depressed (Brittlebank, Scott, William, & Ferrier, 1993), but given how much effect sizes have varied in this literature (see Sumner, Griffith, & Mineka, 2010; Williams et al., 2007, p. 124-126), further research is needed to clarify whether these effects are present across the spectrum of depression severity or whether some protective mediator of these effects exists for undergraduates.

In this study, no data regarding cultural identity of participants was collected and the majority of the sample (57 of 71 participants) identified as female. Depression is more frequently reported by females, so this sample is consistent with samples from much of the research concerning depression, however it must be acknowledged that the findings of my dissertation may be less applicable to men. While there is no precedent in the literature indicating possible relations between MTT, depression and cultural identity, to my knowledge the effect of cultural factors on this association has never been studied; I am therefore not able to exclude the possibility that lack of replication of reduced positive future fluency and Overgeneral AM in this sample could be due to cultural factors.

As noted earlier, methodological choices in this study were guided by the aim of replicating the literature and increasing generalizability of the results to past literature and ongoing research. However, several past studies have used alternative administration and scoring criteria for the AMT. For instance, as noted in the literature review section, Debeer et al. (2009) and Raes et al. (2007) recommend using a version of the AMT with minimal instructions or the SCEPT, an alternative task measuring memory specificity using a sentence completion paradigm to increase sensitivity in non-clinical samples. As can be observed in Figure 8, there was substantial variability in number of specific memories reported in the version of SIT used in this study, but it is possible that the slight negative skew in the distribution of scores masks some true effect on this task. Based on consultation with experts regarding the AMT, as described in the Methods section, the AMT can be scored using a scaled rather than dichotomous response scheme such that responses can be scored as either specific, overgeneral, or intermediate, in which a response does not clearly meet criteria for either or has some level of intermediate specificity. Use of atypical scoring strategies may have increased sensitivity of this task to detect degrees of overgenerality, but it does not explain the lack of replication of results relating to fluency, would not have changed my findings about the relation between the FIT and SIT, and as noted in the Power Analyses section, low sensitivity does not appear to have been the cause of lack of replication in this study. However, it is not possible to rule out the possibility of sampling error.

Future Directions

Although the FIT and SIT were significantly correlated, the magnitude of correlation between conditions was small to medium, indicating that while these tasks share common variance, they also differ. Based on my discussion in the Introduction section, whereas the fluency instruction tasks favour brevity and shallow processing, specificity instruction tasks favour deep processing and are less dependent on speed. Other differences in the quality of responses elicited in each task could include the extent to which each elicits episodic or semantic responses and the richness of these responses. Potential differences relating to these arguably distinct types of memory systems are not reflected in either the task or scoring instructions nor have potential differences on this basis been reported in previous studies. The Autobiographical Interview (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002) was developed to assess specificity of both semantic and episodic components of AM, and could therefore be used in follow-up studies to further explore distinct roles of the episodic and semantic memory systems on performance on the FIT and SIT.

Similarly, potential differences between quality of responses related to either task or group could include temporal distance, or vividness of the memory. In their studies, D'Argembeau and Van der Linden (2004; 2006) reported stronger re-experiencing (or pre-experiencing) of events that were nearer in temporal distance (closer to the present) and that more cohesive episodes were reported for past than future events. Based on these studies, reduced memory specificity in depression could be associated with reduced reported re- or pre-experiencing and reduced specificity could be linked to reporting of memories at greater temporal distances from the present. Future studies could either collect information from participants about the temporal distance and vividness of events reported during FIT or SIT, or employ standard memory or future thinking stimulus to equate for these factors between participants to determine whether temporal distance or vividness mediate or moderate specificity in depressed samples.

As mentioned in the Discussion section, the construct of Time Perspective could play an important role in the MTT or CaR-FA-X Model. Measures of psychopathology have been correlated with the ZTPI, but Time Perspective has not been directly studied alongside the MTT or CaR-FA-X Models, or extensively alongside depression. Specifically, administering this measure alongside other MTT tasks could identify one or more factors mediating MTT differences in depression.

The rationale for the present study was that fluency and specificity instruction tasks are associated with distinct cognitive processes and as a result may be sensitive to different cognitive biases or impairments in depressed samples. Accordingly, as the results highlight, these tasks, though similar, are only moderately correlated and even conditions within each task are not necessarily correlated. However, because the results do not mirror the literature, and because the FIT and SIT do not measure pure cognitive processes and instead approximate underlying ability of several distinct underlying processes based on behavioural task performance, drawing strong conclusions based on the results is difficult. As noted in the literature review, processes implicated in AM and future thinking include executive functioning, construction and elaboration. Other factors that may differ between the FIT and SIT used in this study that were not possible to equate and could account for additional variance between these tasks could include motivation, extent to which tasks elicit semantic or episodic memory content, attention, verbal fluency, self-relevance of stimuli, temporal distance of events reported, and reporting strategy (liberal or conservative criterion).

Although reaction time was recorded in this study, it has not been used as a central dependent variable in previous studies and existing models so reaction time was not examined. Diffusion modeling (Ratcliff, 1978), such as that applied to decision processes in episodic and semantic memory in older adults (Spaniol, Madden, & Voss, 2006) and other investigations in depressed samples (Vallesi, Canalaz, Ballestrieri & Bambilla, 2015; Veiel & Storandt, 2003) could shed light on differences between healthy and depressed samples' performance on fluency and specificity instruction tasks. In a study by Vallesi and colleagues (2015), people with depression modulated their strategies less flexibly based on task instruction to either focus on accuracy or speed in a decision-making task; however in the current study, the DG often performed better than the NDG, and the tasks were counterbalanced, so it is unlikely that inflexible decision-making strategy affected one of the tasks more than the other. Although the MTT tasks used in this study assessed free recall and are therefore not directly translatable to diffusion modeling, the FIT and SIT may differentially emphasize accuracy and speed, so it is important to understand the role of strategy in these tasks. The study by Vallesi and colleagues underscores the importance of exploring potential role of strategy in MTT tasks.

As noted in the introduction, neuroimaging research and cognitive research in healthy samples have contributed significantly to the development of the MTT Model and models of episodic event construction and elaboration. The next logical step in this literature is to translate methodology, such as task variations, and test predictions from these literatures to depressed samples. Recently, several researchers have begun to explore this crossover. Addis, Hach and Tippett (2016) compared performance of healthy and depressed samples on a future-oriented specificity instruction task. The authors included several measures of executive functioning never previously studied alongside past and future event reporting specificity. Addis and colleagues reported that although executive functioning did not differ between the depressed and control samples in their study, semantic clustering as measured by the California Verbal Learning Test was the only significant predictor of future event specificity and did not predict past specificity. Similarly, Anderson and Evans (2015) used a sentence completion task to study past and future thinking in a dysphoric sample. The authors reported that in the Past condition, the dysphoric participants gave overgeneral responses in response to cues of all valence, but in the Future condition, participants only responded in an overgeneral way to emotional words. As illustrated by these two studies, other researchers in the field have recognized the utility and importance of closer integration of the MTT Model into this research. Further studies are needed to study phenomenal characteristics of past and future thinking, such as has been done by D’Argembeau and Van der Linden (2004; 2006) and D’Argembeau et al. (2010).

In their article, Addis and colleagues (2016) noted that their study was conducted in the context of a larger neuroimaging study. Hach, Tippett, and Addis (2014) also examined neural changes in MTT in a depressed sample and noted that while the brain regions activated were similar in both groups, the depressed group demonstrated higher activation of some regions and more interconnectivity of others, which suggests potential differences in the relative importance of subprocesses used to support MTT in people with depression. Given the growing support for a brain-based model of MTT, within the next few years, additional research will emerge examining similarities and differences between healthy and depressed samples on MTT tasks. Specifically, it would be interesting to see whether depressed samples show activation of a characteristic neural network in the first 10 seconds of memory tasks and activation of a different network during the latter half of each trial, a pattern prompting Addis and colleagues (2007) to propose a distinction between Construction and Elaboration processes. In my dissertation, the FIT and SIT were moderately correlated, and as discussed in my Introduction, this could be attributable to the contribution of Construction and Elaboration processes to different degrees in each task. Further, my dissertation failed to replicate findings of group differences on FIT and SIT. Lack of group difference on a behavioural task, such as those used in this study, does not preclude the possibility of different neural responses. Neuroimaging studies could provide a more sensitive method to assess whether depressed samples demonstrate functional differences (regions or magnitudes of neural activity) in Construction or Elaboration processes. Besides increased attention to processes supporting performance in MTT tasks, methodological innovations from neuroimaging literature could include use of less complex tasks to measure specificity and fluency, such as Hassabis et al.'s Episodic Details Task (2007), which has already been used as a more controlled measure of event specificity in healthy samples (D'Argembeau et al., 2010). The clinical implications of neuroimaging studies could include more refined and process-based models of MTT in depression. If functional differences are found between healthy and depressed samples on MTT tasks, use of fMRI before and after treatment targeting reduced reporting specificity in depression could act as a more sensitive treatment outcome measure.

If the current results were considered independently from pre-existing models, which have been applied to fluency and specificity instruction tasks post hoc, very different conclusions about the relationship between depression and MTT would be drawn. For instance, although the differences were not significant, the DG reported more events than did the NDG in most conditions of both tasks. Applying similar post hoc logic, if depression were associated with increased reporting of episodic information, this result could be attributed to a mechanism supporting vulnerability to ruminate in depression, since reporting could be linked to tendency to think about one's past or future or a reflection of well-rehearsed schemas about the self. Alternatively, this could be explained as a sign of failed attempts at problem-solving through generation of new thoughts or ideas. Similarly, although the BDI-II did not show any significant correlations with any conditions of the MTT tasks, there was a significant positive correlation of small magnitude between the Neutral Future condition of the FIT and the BHS, and a small but significant negative correlation between the Positive Future condition of the SIT and the RRS. Neither Neutral conditions in general nor Future conditions of specificity instruction tasks have factored into existing models of MTT in depression, and yet these are the only potential (albeit tenuous) links between MTT and depression detected in this study. Existing models do not suggest any particular reason to predict links between responses to neutral cues and psychopathology. One possible explanation is that people who are hopeless may be more likely to appraise more situations as neutral rather than positive. A higher threshold for identification of positive events would result in a net decrease in subjectively positive experiences, and occurrence of subjectively positive events is theorized to increase mood. The premise that increasing pleasant events increases mood is core to current evidence-based treatments for depression, including Behavioral Activation (Jacobson et al., 1996; Lewinsohn & Graf, 1973) and Mindfulness Based Cognitive Therapy (Williams et al., 2000). Thus, the finding of a positive correlation between Neutral Future fluency would be predicted based on current models of depression.

In attempting to replicate findings from the literature, I conducted ANOVAs of the tasks using only the Negative and Positive conditions of each task because Neutral conditions have not always been included in past research. Using these conditions only, there was a trend toward a Valence by Group interaction for the FIT. The pattern of results was that the DG reported similar numbers of events in the Positive conditions but more Negative events across both Past and Future. This pattern is more in line with a model of increased negative fluency in depression than a model of reduced positive fluency; causal mechanisms could include enhanced rehearsal and strengthening of negative schema through rumination or facilitated retrieval of events because of a mood congruency effect. Inclusion of the Neutral condition in this study however, shows that although there was no significant difference between the DG and NDG in the Neutral task conditions, the pattern was such that the DG reported more events in both the Past and Future Neutral conditions. The pattern of increased reporting in Negative and Neutral conditions indicates that in this sample, had the DG and NDG reported equivalent numbers of events overall, the pattern of results related to Valence would have been similar to the pattern found in the literature. Specifically, if equivalent numbers of events had been reported by both groups, the DG would have reported equivalent numbers of Negative and Neutral events, but fewer Positive events. Overall, this finding suggests that any reduced or exaggerated event reporting based on

Valence in depression is likely to be relative. It may be more useful to compute and report a ratio of Positive to Negative events reported to equate for total events reported in future studies of fluency in depression. Comparing ratios rather than raw numbers of events reported per Valence condition will clarify the relationship between Valence conditions by eliminating the confound of event reporting fluency.

A recent article by Oettingen, Mayer and Portnow (2016) presents an alternative model of the potential role of MTT in depression. In their article, Oettingen and colleagues describe five longitudinal studies demonstrating that positivity of thoughts about the future, described as "fantasies" by the authors, were associated with low levels of depression at the time of assessment but predicted symptoms of depression up to 7 months in the future. Oettingen et al. hypothesize that this relationship is because positivity of fantasies results in decreases in energy and effort toward achievement of positive fantasies. In the study by Oettingen et al., it was the extent of positivity of future thoughts that was measured, rather than specificity or fluency, and the samples used were not depressed, so it is unclear how this finding relates to current depression, but this research highlights the fact that at different stages, depression may not be associated with positive MTT. Research on this phenomenon has also been correlational, so no causal inferences can be drawn. More longitudinal research is needed to clarify whether this finding, reduced positive future fluency and Overgeneral AM are associated with development and/or maintenance of depression.

Another potential confound between fluency and specificity instruction tasks is the subjective interpretation of the valence of cues used. Although the cues were adopted from the literature, and in the case of the SIT, intended valence was validated with valence ratings from psychometric datasets, occasionally during either task, a participant reported an event that was not congruent with the cue's intended valence. The only way to eliminate this potential confound would be for participants to rate the valence of the memories reported in response to each cue, but that would significantly increase testing time and could result in the exclusion of more trials from analysis, resulting in more variability in number of trials per condition. It is also possible that likelihood of reporting events consistent with the intended cue valence is related to depression. Future studies should include participant ratings of trial valence to rule out the possibility that this phenomenon is not random.

In summary, reasons for lack of replication of reduced future fluency and Overgeneral AM in depression in this study could include differences between methodology (including sample, prompting, task administration, and scoring), low clarity and inconsistency of methodology in previous studies, or inaccuracies in existing models and theories of MTT in depression. As summarized above, given the current literature and results, it is not possible to draw strong conclusions about the results in this study, despite the depth and breadth of existing research on this topic. This suggests the need to acknowledge inconsistencies in this literature and address the shortcomings of existing tasks and models before proceeding with development of interventions for depression, such as the MEST.

Chapter 7: Conclusion

The purpose of my dissertation was to answer the question of whether reduced AM specificity and reduced future fluency in depression are linked to the same underlying cognitive process or different processes. Answering this question involved testing predictions regarding fluency and specificity instruction tasks based on the MTT Model, assessing the validity of the MTT Model in a depressed sample, the replicability of the findings of reduced future fluency and reduced memory specificity in depression, and the validity of both the CaR-FA-X and SMS Models. As summarized in the Discussion section, the results of this study challenge basic underlying assumptions of these models. Several informative conclusions can be drawn from my dissertation.

The first conclusion arising from my dissertation concerns the MTT Model. Some predictions based on the MTT Model were supported whereas others were not. The results confirm that past and future conditions of the FIT and SIT followed unique patterns relating to valence: more events were reported in the Neutral and Positive conditions than the Negative condition in the SIT, whereas in the FIT, the fewest events were reported in the Neutral conditions and the most events were reported in the Positive conditions and that these patterns are mirrored across temporal directions in each task. While the pattern of positive correlation across past and future conditions predicted by the MTT Model was found in the FIT, and there were positive correlations between conditions within each temporal direction on the SIT, the past and future conditions of the SIT were not positively correlated as predicted. This suggests that the past and future conditions of the SIT differ either in the cognitive subprocesses supporting them or their relative importance across each temporal direction. This finding neither supports nor refutes the MTT Model because the same processes could still underlie performance in each temporal direction of the SIT, but to a differing degree. However, this study suggests that the validity of the MTT Model is greatly moderated by task demands and underlying processes and that differences and task demands unique to past and future conditions of tasks need to be better understood.

The second conclusion that can be drawn from my dissertation is that the MTT Model also extends to a depressed sample. Expanding on previous studies demonstrating the validity of the MTT Model in a single MTT task in depressed samples (e.g. Macleod et al., 1997), my dissertation demonstrated no difference between depressed and never depressed samples on two tasks assessed concurrently.

The third conclusion arising from my dissertation is that the FIT and SIT, as MTT tasks, share variance, but assess distinct cognitive processes. My dissertation expanded upon work by D'Argembeau and Van der Linden (2004; 2006) demonstrating shared variance between fluency and specificity instruction tasks and expanded upon this literature by demonstrating no differences between healthy and depressed samples.

The fourth and final conclusion suggested by my dissertation is that there are unidentified variables reducing the replicability findings related to MTT in depression. As outlined in the discussion section, unaccounted-for variables in this study and the literature as a whole could include unknown moderators or mediators of the relationship between MTT and depression, inconsistent or unclear methodological documentation, or conceptual issues with fluency and specificity instruction tasks. This is an important and unexpected conclusion given the breadth of research attention dedicated to fluency and specificity in depression. The most obvious potential explanation for lack of replication in this study is recruitment context, but none of the clinical variables available (and many were included) indicated in what way my sample was qualitatively or quantitatively unique. The major implication of this conclusion is that it suggests influence of one or more unknown variables in previous studies assessing the CaR-FA-X Model, identification of which is important to understand to properly evaluate the utility and efficacy of the treatment developed to treat depression based on the CaR-FA-X Model, MEST. Specifically, the lack of replication of the finding of reduced memory specificity in the depressed sample suggests unidentified methodological issues in this literature. Further, as noted in previous sections, scoring of the FIT and SIT in my dissertation highlighted numerous conceptual issues related to MTT that have not been well documented in previous studies that challenge the utility of the SMS model to the literature pertaining to memory specificity; for example, the concern that performance on specificity instruction tasks may be attributable to reporting strategy rather than disruption in episodic event search processes.

In addition to suggesting the four conclusions outlined above, my dissertation makes novel contributions to the field by improving on methodology from past studies. Novel improvements included use of a within-subjects design to study relations between fluency and specificity instruction tasks in depression, inclusion of neutral valence conditions in both tasks, and selection and categorization of cue word valence in the SIT using means from a psycholinguistic database.

In summary, this study illustrates the potential importance of the MTT Model for understanding cognition in depression, and the need to subject the Car-FA-X and MTT Models to further scrutiny. In line with current trends in meta-science (Open Science Collaboration, 2015), my dissertation underscores replication issues in Psychology research in general, and in MTT research in depression specifically. My dissertation, supported by previous research in cognitive neuroscience, cognition, and emerging trends in depression research suggests similarly that cognitive models of depression would benefit from a shift away from distinct models relating to memory and future thinking and toward a broader model of MTT in depression. As noted in the Introduction section of this paper, the conceptualization of memory as a constructive process proved to be highly valuable to the field of memory research, and in challenging models of memory retrieval and accuracy of human memory, it led to important insights into the nature of human memory, such as the identification of areas of the brain responsible for different processes of memory (e.g. Addis, Wong, & Schachter, 2007). A growing body of research supports the eventual convergence of the MTT and depression literatures and development of a more nuanced model of MTT in depression may similarly yield important insights, such as transfer of tasks and analytic methods from cognitive research to depressed samples and use of neuroimaging to assess MTT processes implicated in depression to answer the question of whether or how MTT differs between depressed and non-depressed samples.

Appendix A: A priori group criteria and characteristics of original NDG, Intermediate and DG

In total, 143 participants took part in this study. Of these participants, only a subset met inclusion criteria based on the a priori criteria. Using BDI-II score greater than 13 and endorsement of criteria indicative of current MDE on the MINI, 23 participants met criteria for inclusion in the DG; 63 participants had BDI-II scores below 14 and did not endorse any history of depression on the MINI, meeting a priori criteria for the NDG; 56 participants had either a BDI-II score above 13 or endorsed either a current or past MDE on the MINI and were classified as an Intermediate Group; one participant did not complete the MINI and therefore could not be included in the analyses. Following initial examination of the data, I noted that the intermediate group results did not relate linearly with the results of the DG and NDG across MTT conditions, likely due to sample heterogeneity within this intermediate sample on a variety of variables. For simplicity of interpretation and to increase power, this group was excluded from the main analyses. In addition, the range of BAI and BHS scores in the NDG was of concern, since both anxiety and suicidality have been linked with MTT processes.

To reduce possible confounds and increase sensitivity, following initial screening of the data suggesting possible mental health concerns in the NDG, more stringent group criteria were adopted.

Characteristics of Never Depressed Group using original group criteria

	N	Minimum	Maximum	Mean	Std. Deviation
Beck Depression Inventory Score	63	.0	13.0	5.365	3.9931
Participant Age	62	17	47	20.19	5.254
FAS Raw Score	63	17.0	56.0	33.381	8.0431
Animals Raw Score	63	10.0	29.0	19.825	4.4669
Longest Digit Span Forward Score	62	4.0	9.0	6.242	1.1407
Digit Span Forward Score	62	4.0	15.0	9.565	2.2296

Wide Range Achievement Test Raw Score	63	46.0	69.0	57.651	5.5537
Beck Anxiety Inventory Score	62	.0	32.0	7.677	6.6574
Beck Hopelessness Scale Score	62	.0	6.0	1.968	1.6293
Ruminative Response Scale Score	61	20.0	57.0	34.115	9.1289
Trauma History Screen Total Events Reported	62	.0	11.0	1.661	1.9497
Trauma History Screen Total Number of Distressing Events	62	.0	7.0	.613	1.2194

Characteristics of Intermediate Group using original group criteria

	N	Minimum	Maximum	Mean	Std. Deviation
Beck Depression Inventory Score	53	2.0	32.0	13.755	7.3273
Participant Age	53	17.0	43	20.21	5.336
FAS Raw Score	51	15.0	64.0	36.784	11.2558
Animals Raw Score	51	10.0	34.0	20.784	4.9287
Longest Digit Span Forward Score	52	3.0	9.0	6.731	1.3736
Digit Span Forward Score	52	4.0	14.0	10.077	2.2215
Wide Range Achievement Test Raw Score	51	47.0	69.0	58.784	5.2357
Beck Anxiety Inventory Score	51	2.0	38.0	16.471	9.7905
Beck Hopelessness Scale Score	50	.0	15.0	4.360	4.0394
Ruminative Response Scale Score	49	25.0	59.0	41.449	9.9876
Trauma History Screen Total Events Reported	49	.0	30.0	5.510	6.1309
Trauma History Screen Total Number of Distressing Events	49	.0	5.0	1.469	1.4157

Characteristics of Depressed Group using original group criteria

	N	Minimum	Maximum	Mean	Std. Deviation
Beck Depression Inventory Score	23	14.0	43.0	25.304	8.2265
Participant Age	23	17	41	20.96	6.428
FAS Raw Score	23	22.0	57.0	38.304	9.0576
Animals Raw Score	23	12.0	28.0	21.043	3.5481
Longest Digit Span Forward Score	21	4.0	9.0	6.429	1.4687
Digit Span Forward Score	20	7.0	14.0	9.850	2.0072
Wide Range Achievement Test Raw Score	20	55.0	67.0	60.700	3.6143
Beck Anxiety Inventory Score	23	.0	45.0	23.478	10.9040
Beck Hopelessness Scale Score	21	.0	69.0	10.238	14.3349
Ruminative Response Scale Score	20	8.0	77.0	51.750	16.6919
Trauma History Screen Total Events Reported	20	.0	109.0	11.450	23.7120
Trauma History Screen Total Number of Distressing Events	20	.0	8.0	1.900	1.9974

Appendix B: Administration and Scoring Criteria for the FIT
Fluency Instruction Task (FIT)

Subjects are given three future time periods (the next week, the next year, the next five to ten years) and three past time periods (the past week, the past year, the past five to ten years) and asked to try to think of positive things (things they are looking forward to), negative things (things they are not looking forward to) and neutral things (neither looking forward to or not looking forward to) for each of those time periods.

The order of presentation of neutral, negative and positive temporal direction (past or future) conditions should be counterbalanced across subjects, although within each condition the time periods are always presented in the same order (week, year, 5-10 years). Verbal instructions to participants are italicized.

If subject says during the thinking time that they can't think of anything or, for example, that there is nothing that they are looking forward to over the next week, say "*that's OK, but just keep trying to think until I tell you to stop"*.

Instruction script for FIT:

"Now I'd like to ask you to think about things that have happened to you in the past or might happen to you in the future. I will give you different time periods either in the future or past, one at a time, and I'd like you to try to think of things that might happen to you or did happen to you in those time periods. I will give you a minute to try to think of as many things as you can. It doesn't matter whether the things are trivial or important, just say what comes to mind. But, they should be things have happened or that you think will definitely happen or are at least quite likely to happen. If you can't think of anything or if you can't think of many things, that's fine, but just keep trying until the time limit is up."

POSITIVE FUTURE:

"First I'm going to ask you to think of positive things in the future. So, I'd like you to try to think of things that you are looking forward to, in other words, things that you will enjoy. So, I want you to give me as many things as you can that you're looking forward to over the next week including today".

(Researcher gives one minute and audio records and writes down as close to verbatim as time allows what subject says)

Now, I'd like you to do the same but this time I want you to give me things that you're looking forward to over the next week.

(Researcher does same as for one week)

Now, I'd like you to do the same but this time I want you to give me things that you're looking forward to over the next five to ten years.

(Researcher does same as for previous)

NEGATIVE FUTURE:

"Now, I'd like you to think of things that you're worried about or not looking forward to, in other words, things that you would rather not be the case or rather not happen. So, I want you to give me as many things as you can that you're worried about or not looking forward to over the next week including today".

(Researcher does same as for previous)

"Now I want you to give me as many things as you can that you're worried about or not looking forward to over the next year"

(Researcher does same as for previous)

Finally, I want you to give me as many things as you can that you're worried about or not looking forward to over the next five to ten years"

(Researcher does same as for previous)

POSITIVE PAST:

"Now I'm going to ask you to think of positive things in the past. So, I'd like you to try to think of things that you enjoyed. So, I want you to give me as many things as you can that you enjoyed over the past week including today".

(Researcher gives one minute and audio records and writes down as close to verbatim as time allows what subject says)

Now, I'd like you to do the same but this time I want you to give me things that you've enjoyed in the past year.

(Researcher does same as for one week)

Now, I'd like you to do the same but this time I want you to give me things that you've enjoyed in the past 5-10 years.

(Researcher does same as for previous)

NEGATIVE PAST:

"Now, I'd like you to think of things that you did not enjoy or would rather not have happened. So, I want you to give me as many things as you can that did not enjoy or would rather not have happened in the past week including today".

(Researcher does same as for previous)

"Now I want you to give me as many things as you can that did not enjoy or would rather not have happened over the past year"

(Researcher does same as for previous)

"Now I want you to give me as many things as you can that did not enjoy or would rather not have happened over the past 5-10 years"

(Researcher does same as for previous)

NEUTRAL PAST:

"Now I'm going to ask you to think of emotionally neutral things in the past. So, I'd like you to try to think of things that happened to you in the past that you do not have strong feelings about either way. So, I want you to give me as many things as you can over the past week including today".

(Researcher gives one minute and audio records and writes down as close to verbatim as time allows what subject says)

Now, I'd like you to do the same but this time I want you to give me things that have happened in the past year.

(Researcher does same as for one week)

Now, I'd like you to do the same but this time I want you to give me things that have happened in the past 5-10 years.

(Researcher does same as for previous)

NEUTRAL FUTURE:

"Now I'm going to ask you to think of emotionally neutral things in the future. So, I'd like you to try to think of things that could happen to you in the future that you do not have strong feelings about either way. So, I want you to give me as many things as you can over the next week including today".

(Researcher does same as for previous)

"Now I want you to give me as many things that could happen to you in the future that you do not have strong feelings about over the next year"

(Researcher does same as for previous)

Finally, I want you to give me as many things that could happen to you in the future that you do not have strong feelings about over the next five to ten years"

(Researcher does same as for previous)

Scoring

The score is the total number of responses given in a particular condition time period. Thus, there will be a number for each of the categories (3 time periods x 3 valence conditions x 2 temporal directions). Where a subject repeats a response across different time categories only include it the first time it is mentioned.

Appendix C: Administration and Scoring Criteria for the SIT

Specificity Instruction Task

Administration notes, which will not be read aloud to participants are italicized.

Now I am going to ask you to describe specific personal memories and future events. I will read aloud and show you words on a card one at a time and ask you to tell me a specific event from either your personal past or future related to that word. While you respond, I will be audio-recording and writing down your answers. A specific event is an event is **an event that lasted less than a day, and occurred at a particular time and place**. So if I said the word “good” – it would not be OK to say, “I always enjoy a good party”, because that does not mention a specific event. But it would be OK to say “I had a good time at Jane’s party” because that is a specific event. The event that you report be something that happened in the near past or future or a long time in the past or future, and that it can be an important event or trivial event but that the event should be of something that happens or happened at a particular time on a particular day. Events cannot have occurred or be imagined to occur within 7 days of today. It is important to retrieve a different event for each cue word. Do you have any questions? *[Instructions will be clarified as needed]*

Before we begin, you will do two practice words to get the hang of it. Are you ready to try?

Can you tell me one specific moment or event from your past that the word *relieved* reminds you of?’

Subject is given 60s to respond for each word.

Can you tell me one specific moment or event from your future that the word *tired* reminds you of?’

Subject is given 60s to respond for each word.

If the type of event that the participants report is not specific, or if participants retrieved the same event to more than one cue or offered responses that related to the opposite temporal orientation, they will prompted using the most appropriate phrase below after it is clear that they have finished describing the event: “Can you think of a specific time—one particular episode?” “What is the event that you are thinking of there?”’ or “Can you tell me a bit more about that event?”. Participants will be given feedback on practice trials and corrected if specific events are not reported. Once the participant has reported specific events for the words above, the experimental trials will begin.

Great, now you know how to do the task. Let’s begin the study task. For the first set of words, I will ask you to report events from your past.

Can you tell me one specific moment or event from your past that the word *X* reminds you of?’

Can you tell me one specific moment or event from your future that the word *X* reminds you of?’

Stimuli:

The words will be presented in random order; the lists will be counterbalanced such that some participants report past events for list A and future events for list B and vice versa. The instructions will be counterbalanced randomly so that half of the participants report past events first and half report future events. Cue words will be presented on 12.5 cm _ 7.5 cm laminated cards and were written in black ink in capital letters 3.5 cm high.

List A:
COMPLIMENT
GIFT
LAUGH
SAFE
HAPPY
NERVOUS
FAILURE
GUILT
DANGER
BLAME
ADVICE
SHOP
OBEDIENCE
UNCLE
INTEREST

List B:
SURPRISE
SUCCESS
SMILE
TRAVEL
PROUD
SAD
GRIEF
HURT
TEAR
LATE
PACKAGE
FASHION
LIBRARY
OCCASION
WALK

The cumulative time taken to give all subsequent responses will be recorded, the prompting procedure being repeated if these responses remained inappropriately general. If subjects did not retrieve any specific memory in the time available, a time of 60 s should be recorded, and the experimenter will proceed to the next item.

Scoring instructions: events will be coded specific if they referred to a particular event that occurred within the course of 1 day (e.g. ''at John's party two weeks ago''). Non-specific events will be qualified as either categoric (a memory that summarises a number or category of events; e.g., ''taking a bath''), extended (a memory of a period lasting longer than 1 day; e.g., ''last summer holidays''), or semantic associates (verbal association to the cue or thought about the future; e.g., ''my mother''). Failures to provide a memory will be classified as omissions. Finally, there is an Other category that included all incomplete responses and all responses for which the instructions had not been followed (e.g., memory not older than 7 days; referring to an event already mentioned).

Appendix D: Additional FIT and SIT scoring guidelines including rules of thumb generated by two research assistants and I; rules of thumb and scoring in general were informed by unpublished documentation provided by Drs. Mark Williams (2015), Tim Dalgleish (2015), and Andrew Macleod (2014) or by associates of their labs on their behalf.

Scoring Rules of Thumb

Purpose: I created this document with two research assistants while training to score the FIT and SIT.

Please use this document to make notes about ambiguous scoring scenarios to discuss as a group or query the creators

Here are some that I have noticed so far and my thoughts about consistency management:

- repeats from practice trials
  - do not count as repeats
- repeats across temporal directions
  - do not count as repeats
- repeats across valence
  - do NOT count as repeats

- For SIT and FIT: if someone gives a response for the wrong temporal direction, count if they don't mention if for the correct temporal direction, but exclude if they repeat it; they should receive a prompt for SIT to generate a different event for the correct temporal direction
- For SIT: to distinguish categorical from specific responses, allow 10 seconds after a prompt for the person to elaborate, and if they do not elaborate or repeat the same info, count as categorical (in early trials I gave more time than this)

(FIT)
Negative valence - Not doing something
Positive valence - Doing something
Is this a repeat?
Example:
Future Pos - Getting into program
Future Neg - Not getting into program

→ NOT COUNTED AS REPEAT

FIT: if someone says "all the things listed in previous condition", it counts as a repeat. e.g. futneutral1week: class, subway, dinner; futneutral1year: same as above.

For past & future (In SIT), if multiple types of events are listed (such as using words as "like," "probably," etc.), instead of a concrete specific event, it may not be specific

→ IF AN EVENT IN THE FUTURE WOULD BE SPECIFIC IN THE PAST, THEN THAT FUTURE EVENT IS SPECIFIC

SIT: if a participant gives an INCORRECT FIRST RESPONSE (<7 days, wrong temporal direction), start counting time after the corrective prompt is given. Record the prompt in #prompts; count prompts given after the participant's event in Total prompts and do not include #prompts in total prompts
SIT: If the person REPEATS an event, count it as a new variable "repeat" (6) in SPSS; you may need to add this variable to the database

- In this case, if Maddy prompted to come up with a different event after the repeat, do not count that in #prompts.

SIT: if someone self-corrects their response, or changes the event without a prompt, the latency time is when the participant starts reporting the SCORED event - the one they give as their last answer.

For SIT: If it is an event that reoccurs, they need to give more than just a date and time, but need to provide enough details and unique descriptions to distinguish it from every other event.

For SIT: If in doubt between categorical and specific, give benefit of the doubt.

For SIT: pretend it is specific from the start and try to see if the details are all specific. Assume it's specific and see if it sticks.

For SIT: ask what details someone would need to provide to ensure the event is categorized as specific and ask if it is reasonable to expect that they would give that based on task instructions.

For SIT: if they give a future event that could be categorical, they need to attach a detailed timeframe for the event to be specific (ie. not just 'lunch with my uncle on a Sunday in Toronto, because that could be any time in a category)

For future SIT: if a categorical event like a birthday, need to come up with at least one clear unique detail distinguishing this event from a previous similar event to get specific response

Distinguishing between categoric and specifc - is it a routine? do you have to go out of your way to schedule it? For example, getting a physical (participant 494 - shop)

Participant 494 - blame (~1.04.00) - categorical or specific. "On January 1, 2016, I will go to a party with my friend Bob, as we usually do, and we will talk about the past year and I will blame myself for mental issues, or whatever has gone wrong in the past year. Usually Bob and I go out on New Year's and reflect on the past year."
THE ABOVE IS A SPECIFIC EVENT

Appendix E: Permissions from publishers to reproduce Figures 1-3

**CAMBRIDGE UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS**

Dec 13, 2016

This Agreement between Madelaine Burley ("You") and Cambridge University Press ("Cambridge University Press") consists of your license details and the terms and conditions provided by Cambridge University Press and Copyright Clearance Center.

License Number	3855421197321
License date	Apr 24, 2016
Licensed Content Publisher	Cambridge University Press
Licensed Content Publication	Behavioral and Brain Sciences
Licensed Content Title	The evolution of foresight: What is mental time travel, and is it unique to humans?
Licensed Content Author	Thomas Suddendorf and Michael C. Corballis
Licensed Content Date	Oct 29, 2007
Licensed Content Volume Number	30
Licensed Content Issue Number	03
Start page	299
End page	313
Type of Use	Dissertation/Thesis
Requestor type	Not-for-profit
Portion	Text extract
Number of pages requested	1
Order reference number	
Territory for reuse	World
Title of your thesis / dissertation	Disentangling Cue Valence, Temporal Orientation and Task Instructions During Mental Time Travel In Depression
Expected completion date	Aug 2016
Estimated size(pages)	150
Requestor Location	Madelaine Burley Department of Psychology Ryerson University 350 Victoria Ave. Toronto, ON M5B 2K3 Canada Attn: Madelaine Burley
Billing Type	Invoice
Billing Address	Madelaine Burley Department of Psychology Ryerson University 350 Victoria Ave. Toronto, ON M5B 2K3 Canada Attn: Madelaine Burley

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 13, 2016

This Agreement between Madelaine Burley ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3852810809013
License date	Apr 19, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Memory and Language
Licensed Content Title	Memory and the self
Licensed Content Author	Martin A. Conway
Licensed Content Date	October 2005
Licensed Content Volume Number	53
Licensed Content Issue Number	4
Start Page	594
End Page	628
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 5
Title of your thesis/dissertation	Disentangling Cue Valence, Temporal Orientation and Task Instructions During Mental Time Travel In Depression
Expected completion date	Aug 2016
Estimated size (number of pages)	150
Elsevier VAT number	GB 494 6272 12
Requestor Location	Madelaine Burley Department of Psychology Ryerson University 350 Victoria Ave. Toronto, ON M5B 2K3 Canada Attn: Madelaine Burley
Billing Type	Invoice
Billing Address	

AMERICAN PSYCHOLOGICAL ASSOCIATION

Title: Autobiographical memory specificity and emotional disorder.

Author: Williams, J. Mark G.; Barnhofer, Thorsten; Crane, Catherine; Herman, Dirk; Raes, Filip; Watkins, Ed; Dalgleish, Tim

Publication: Psychological Bulletin

Publisher: American Psychological Association

Date: Jan 1, 2007

Copyright © 2007, American Psychological Association

Logged in as:
Madelaine Burley
Account #:
3001020713
LOGOUT

Grant Reuse

APA hereby grants permission at no charge for the following material to be reused according to your request, subject to a required credit line. Author permission is not required in this instance.
• Single text excerpts of less than 400 words (or a series of text excerpts that total less than 800 words) from APA books and journals.
• 1-3 Chart/Graph/Tables or Figures from 1 article or chapter.
• Note that scales, measures, instruments, questionnaires, photographs, or creative images are NOT included in this gratis reuse.
• Also, the abstract of a journal article may not be placed in a database for subsequent redistribution without contacting APA for permission.

BACK | CLOSE WINDOW

References

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

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, *47*(11), 2222-2238. doi: 10.1016/j.neuropsychologia.2008.10.026

Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, *18*(2), 227-237. doi: 10.1002/hipo.20405

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363-1377. doi: 10.1016/j.neuropsychologia.2006.10.016

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science*, *19*(1), 33-41. doi: 10.1111/j.14679280.2008.02043.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association. (2013, June 6). Highlights of changes from DSM-IV-TR to DSM-5. Retrieved from: http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf

Anderson, R. J., & Evans, G. L. (2015). Mental time travel in dysphoria: Differences in the content and subjective experience of past and future episodes. *Consciousness and Cognition, 37*, 237-248. doi: 10.1080/09658211.2015.1046134

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., ... & Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *The Annals of Family Medicine*, *8*(4), 348-353. doi: 10.1370/afm.1139

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences, 5*(12), 533-539. doi: http://dx.doi.org/10.1016/S1364-6613(00)01804-0

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145-159. doi:10.1080/13854040590947362

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893-897. doi:10.1037/0022-006X.56.6.893

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.

Beck, A. T., & Steer, R. A. (1988). *Beck Hopelessness Scale.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561-571. doi:10.1001/archpsyc.1961.01710120031004.

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism:

The hopelessness scale. *Journal of Consulting and Clinical Psychology*, *42*(6), 861-865. doi:10.1037/h0037562

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49-57. doi: 10.1016/j.tics.2006.11.004

Carlson, E. B., Smith, S. R., Palmieri, P. A., Dalenberg, C., Ruzek, J. I., Kimerling, R., ... & Spain, D. A. (2011). Development and validation of a brief self-report measure of trauma exposure: The Trauma History Screen. *Psychological Assessment*, *23*(2), 463-477. doi: 10.1037/a0022294

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition.* Hillsdale, N.J.: Lawrence Erlbaum.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*(4), 594-628. doi:10.1016/j.jml.2005.08.005

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261. doi: 10.1037/0033-295X.107.2.261

D'Argembeau, A., Ortoleva, C., Jumentier, S., & Van der Linden, M. (2010). Component processes underlying future thinking. *Memory & Cognition*, *38*(6), 809-819. doi: 10.3758/MC.38.6.809

D’Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, *13*(4), 844-858. doi: 10.1016/j.concog.2004.07.007

D'Argembeau, A., & Van der Linden, M. (2005). Influence of emotion on memory for temporal information. *Emotion*, *5*(4), 503. doi: 10.1037/1528-3542.5.4.503

D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition*, *15*(2), 342-350. doi: 10.1016/j.concog.2005.09.001

Dalgleish, T., Williams, J. M. G., Golden, A. M. J., Perkins, N., Barrett, L. F., Barnard, P. J., ... & Watkins, E. (2007). Reduced specificity of autobiographical memory and depression: The role of executive control. *Journal of Experimental Psychology: General*, *136*(1), 23-42. doi: 10.1037/0096-3445.136.1.23

Debeer, E., Hermans, D., & Raes, F. (2009). Associations between components of rumination and autobiographical memory specificity as measured by a Minimal Instructions Autobiographical Memory Test. *Memory*, *17*(8), 892-903. doi: 10.1080/09658210903376243

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

Dozois, D. J., & Covin, R. (2004). The Beck Depression Inventory-II (BDI-II), Beck Hopelessness Scale (BHS), and Beck Scale for Suicide Ideation (BSS). In Hilsenroth, M. J. & Segal, D. (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2: Personality assessment* (pp. 50-69). Hoboken, NJ, US: John Wiley & Sons Inc.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Garcia-Toro, M., & Aguirre, I. (2007). Biopsychosocial model in depression revisited. *Medical Hypotheses*, *68*(3), 683-691. doi: 10.1016/j.mehy.2006.02.049

Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., ... & Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory*, *17*(6), 609-623. doi: 10.1080/09658210902939348

Griffith, J. W., Sumner, J. A., Raes, F., Barnhofer, T., Debeer, E., & Hermans, D. (2012). Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, S21-S31. doi: 10.1016/j.jbtep.2011.05.008

Hach, S., Tippett, L. J., & Addis, D. R. (2014). Neural changes associated with the generation of specific past and future events in depression. *Neuropsychologia*, *65*, 41-55. doi: 10.1016/j.neuropsychologia.2014.10.003

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*(7), 299-306. doi: http://dx.doi.org/10.1016/j.tics.2007.05.001

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest*, *3*(2), 39-77.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., ... & Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*(2), 295. doi: http://dx.doi.org/10.1037/0022-006X.64.2.295

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy*

And Research, *11*(3), 289-299. doi: 10.1007/BF01186280

Keough, K. A., Zimbardo, P. G., & Boyd, J. N. (1999). Who's smoking, drinking, and using drugs? Time perspective as a predictor of substance use. *Basic and Applied Social Psychology*, *21*(2), 149-164. doi: 10.1207/S15324834BA210207

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care*, *41*(11), 1284-1292. Retrieved from http://www.jstor.org/stable/3768417 .

Lavender, A., & Watkins, E. (2004). Rumination and future thinking in depression. *British Journal of Clinical Psychology*, *43*(2), 129-142. doi: 10.1348/014466504323088015

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychology and Aging*, *17*(4), 677. doi: http://dx.doi.org/10.1037/0882-7974.17.4.677

Lewinsohn, P. M., & Graf, M. (1973). Pleasant activities and depression. *Journal of Consulting and Clinical Psychology*, *41*(2), 261. doi: http://dx.doi.org/10.1037/h0035142

Luong, N., Burley, M. C. & Girard, T. A. (in preparation). Cognitive components of future thinking.

MacLeod, A. K., & Byrne, A. (1996). Anxiety, depression, and the anticipation of future positive and negative experiences. *Journal of Abnormal Psychology*,*105*(2), 286-289. doi: 10.1037/0021-843X.105.2.286

MacLeod, A. K., Rose, G. S., & Williams, J. M. G. (1993). Components of hopelessness about the future in parasuicide. *Cognitive Therapy and Research*, *17*(5), 441-455. doi: 10.1007/BF01173056

MacLeod, A. K., & Salaminiou, E. (2001). Reduced positive future-thinking in depression:

Cognitive and affective factors. *Cognition & Emotion*, *15*(1), 99-107. doi:10.1080/0269993004200006/

MacLeod, A. K., Tata, P., Kentish, J., & Jacobsen, H. (1997). Retrospective and prospective cognitions in anxiety and depression. *Cognition & Emotion*,*11*(4), 467-479. doi: 10.1080/026999397379881

Medical Research Council. A Randomised controlled clinical trial of Memory Specificity Training (MEST) for depression. In ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000- [cited 2014 Apr 15]. Retrieved from http://clinicaltrials.gov/show/NCT01882452 NLM Identifier: NCT01882452.

Moore, R. G., Watts, F. N., & Williams, J. M. G. (1988). The specificity of personal memories in depression. *British Journal of Clinical Psychology*, *27*(3), 275-276. doi: 10.1111/j.2044-8260.1988.tb00790.x

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J.,Coryell, W. . . . Maser, J. D. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *American Journal of Psychiatry*, *156*, 1000 –1006. Retrieved from http://journals.psychiatryonline.org/article.aspx?articleid=173561

Neshat-Doost, H. T., Dalgleish, T., Yule, W., Kalantari, M., Ahmadi, S. J., Dyregrov, A., & Jobson, L. (2013). Enhancing autobiographical memory specificity through cognitive training: An intervention for depression translated from basic science. *Clinical Psychological Science*, *1*(1), 84-92. doi: 10.1177/2167702612454613

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400-424. doi: 10.1111/j.17456924.2008.00088.x

Oettingen, G., Mayer, D., & Portnow, S. (2016). Pleasure Now, Pain Later Positive Fantasies About the Future Predict Symptoms of Depression. *Psychological Science, 27*(3), 345-353. doi: 10.1177/0956797615620783

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi: 10.1126/science.aac4716

Park, R. J., Goodyer, I. M., & Teasdale, J. D. (2002). Categoric overgeneral autobiographical memory in adolescents with major depressive disorder. *Psychological Medicine*, *32*(2), 267-276. doi: http://dx.doi.org/10.1017/S0033291701005189

Pearson, C., Janz, T. & Ali, J. (2013, September 18). Mental and substance use disorders in Canada, *Health at a Glance, September*. Statistics Canada Catalogue no. 82-624-X. Retrieved on October 21, 2013 from: http://www.statcan.gc.ca/pub/82-624-x/2013001/article/11855-eng.htm

Piolino, P., Desgranges, B., & Eustache, F. (2009). Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia*, *47*(11), 2314-2329. doi: 10.1016/j.neuropsychologia.2009.01.020

Raes, F., Hermans, D., Williams, J. M. G., Demyttenaere, K., Sabbe, B., Pieters, G., & Eelen, P. (2006). Is overgeneral autobiographical memory an isolated memory phenomenon in major depression?. *Memory, 14*(5), 584-594. doi: 10.1080/09658210600624614

Raes, F., Hermans, D., Williams, J. M. G., & Eelen, P. (2007). A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. *Memory*, *15*(5), 495-507. doi: 10.1080/09658210701390982

Raes, F., Williams, J. M. G., & Hermans, D. (2009). Reducing cognitive vulnerability to depression: A preliminary investigation of MEmory Specificity Training (MEST) in inpatients with depressive symptomatology. *Journal of Behavior Therapy and Experimental Psychiatry, 40*(1), 24-38. doi: 10.1016/j.jbtep.2008.03.001

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676-682. doi: 10.1073/pnas.98.2.676

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59. doi: http://dx.doi.org/10.1037/0033-295X.85.2.59

Rawal, A., & Rice, F. (2012). Examining overgeneral autobiographical memory as a risk factor for adolescent depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(5), 518-527. doi: 10.1016/j.jaac.2012.02.025

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, *17*(2), 265-279. doi: http://dx.doi.org/10.1080/10705511003659417

Reisberg, D. E., & Hertel, P. E. (2004). *Memory and emotion*. New York, NY: Oxford University Press.

Rush, A., Trivedi, M., Wisniewski, S., Nierenberg, A., Stewart, J., Warden, D., ... & Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR* D report. *American Journal of Psychiatry*, *163*(11), 1905-1917. doi:10.1176/appi.ajp.163.11.1905

Sarkohi, A., Bjärehed, J., & Andersson, G. (2011). Links between future thinking and autobiographical memory specificity in major depression. *Psychology*, *2*(3), 260-264. doi:10.4236/psych.2011.230

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182. doi: 10.1037/0003-066X.54.3.182

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers.* New York, NY: Houghton Mifflin Company.

Schacter, D. L., & Addis, D. R. (2007). Constructive memory: The ghosts of past and future. *Nature, 445* (7123), 27-27. doi:10.1038/445027a

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*(4), 677-694. doi: 10.1016/j.neuron.2012.11.001

Segal, Z. V. (1988). Appraisal of the self-schema construct in cognitive models of depression. *Psychological Bulletin*, *103*(2), 147. doi: 10.1037/0033-2909.103.2.147

Serrano, J. P., Latorre, J. M., Gatz, M., & Montanes, J. (2004). Life review therapy using autobiographical retrieval practice for older adults with depressive symptomatology. *Psychology and Aging*, *19*(2), 272. doi: 10.1037/0882-7974.19.2.272

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Supplement 20), 22–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9881538

Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T. . . . Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry*, *157*, 229 –233. doi:10.1176/appi.ajp.157.2.229

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101-117. doi: 10.1037/0278-7393.32.1.101

Spreng, R. N., & Levine, B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Memory & Cognition*, *34*(8), 1644-1651. Retrieved from http://link.springer.com/article/10.3758/BF03195927#page-1

Stange, J. P., Hamlat, E. J., Hamilton, J. L., Abramson, L. Y., & Alloy, L. B. (2013). Overgeneral autobiographical memory, emotional maltreatment, and depressive symptoms in adolescence: Evidence of a cognitive vulnerability–stress interaction. *Journal of Adolescence*, *36*(1), 201-208. doi: 10.1016/j.adolescence.2012.11.001

Steer, R. A., Cavalieri, T. A., Leonard, D. M., & Beck, A. T. (1999). Use of the Beck Depression Inventory for primary care to screen for major depression disorders. *General Hospital Psychiatry*, *21*(2), 106-111. doi: 10.1016/S0163-8343(98)00070-X

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory—Second Edition in a sample of college students. *Depression and Anxiety*, *19*(3), 187-189. doi: 10.1002/da.20002

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*,*123*(2), 133-167. Retrieved from http://cogprints.org/725/1/MentalTimeTravel.txt

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(03), 299-313. doi: http://dx.doi.org/10.1017/S0140525X07001975

Sumner, J. A. (2012). The mechanisms underlying overgeneral autobiographical memory: An evaluative review of evidence for the CaR-FA-X model. *Clinical Psychology Review*, *32*(1), 34-48. doi:10.1016/j.cpr.2011.10.003

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy*, *48*(7), 614-625. doi: 10.1016/j.brat.2010.03.013

Sumner, J. A., Mineka, S., Zinbarg, R. E., Craske, M. G., Vrshek-Schallhorn, S., & Epstein, A. (2014). Examining the long-term stability of overgeneral autobiographical memory. *Memory*, *22*(3), 163-170. doi: 10.1080/09658211.2013.774021

Svanborg, P., & Åsberg, M. (2001). A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Åsberg Depression Rating Scale (MADRS). *Journal of Affective Disorders*,*64*(2), 203-216. DOI: 10.1016/S0165-0327(00)00242-1

Szpunar, K. K. (2010). Episodic future thought: An emerging concept. *Perspectives on Psychological Science*, *5*(2), 142-162. doi: 10.1177/1745691610362350

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*,*104*(2), 642-647. doi: 10.1073/pnas.0610082104

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167-177. doi: 10.1016/S0887-6177(97)00095-4

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*,*27*(3), 247-259. doi:

10.1023/A:1023910315561

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

Vallesi, A., Canalaz, F., Balestrieri, M., & Brambilla, P. (2015). Modulating speed-accuracy strategies in major depression. *Journal of Psychiatric Research*, *60*, 103-108. doi:10.1016/j.jpsychires.2014.09.017

van Vreeswijk, M. F., & de Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy*, *42*(6), 731-743. doi:10.1016/S0005-7967(03)00194-3

Veiel, L. L., & Storandt, M. (2003). Processing costs of semantic and episodic retrieval in younger and older adults. *Aging, Neuropsychology, and Cognition*, *10*(1), 61-73. doi: 10.1076/anec.10.1.61.13458

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191-1207. doi: 10.3758/s13428-012-0314-x

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Weiler, J. A., Suchan, B., & Daum, I. (2010). Foreseeing the future: occurrence probability of imagined future events modulates hippocampal activation. *Hippocampus*, *20*(6), 685-690. doi: 10.1002/hipo.20695

Wessel, I., Merckelbach, H., & Dekkers, T. (2002). Autobiographical memory specificity, intrusive memory, and general memory skills in Dutch–Indonesian survivors of the

World War II era. *Journal of Traumatic Stress*, *15*(3), 227-234. doi: 10.1023/A:1015207428675

Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory—Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology*, *56*(4), 545-551. doi: 10.1002/(SICI)1097-4679(200004)56:4<545::AID-JCLP7>3.0.CO;2-U

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test--Fourth Edition*. Lutz, FL: Psychological Assessment Resources.

Wisco, B. E., & Nolen-Hoeksema, S. (2010). Valence of autobiographical memories: The role of mood, cognitive reappraisal, and suppression. *Behaviour Research and Therapy*, *48*(4), 335-340. doi: 10.1016/j.brat.2009.11.009

World Health Organization (2004). *Global burden of disease.* Retrieved from http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf .

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*(1), 122. doi: 10.1037/0033-2909.133.1.122

Williams, J. M., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology*, *95*(2), 144. doi:10.1037/0021-843X.95.2.144

Williams, J. M. G., & Dritschel, B. H. (1988). Emotional disturbance and the specificity of autobiographical memory. *Cognition & Emotion*, *2*(3), 221-234. doi: 10.1080/02699938808410925

Williams, J. M. G., Ellis, N. C., Tyers, C., Healy, H., Rose, G., & MacLeod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Memory & Cognition*, *24*(1), 116-125. doi: 10.3758/BF03197278

Williams, J. M. G., & Scott, J. (1988). Autobiographical memory in depression. *Psychological Medicine*, *18*(03), 689-695. doi: http://dx.doi.org/10.1017/S0033291700008370

Williams, J. M. G., Teasdale, J. D., Segal, Z. V., & Soulsby, J. (2000). Mindfulness-based cognitive therapy reduces overgeneral autobiographical memory in formerly depressed patients. *Journal of Abnormal Psychology*, *109*(1), 150-155. doi: 10.1037/0021-843X.109.1.150

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6-10. Retrieved from http://websites.psychology.uwa.edu.au/school/MRCDatabase/MDWilsonMRC.pdf

Young, K. D., Bellgowan, S. F., Bodurka, J., & Drevets, W. C. (2013). Behavioral and neurophysiological correlates of autobiographical memory deficits in patients with depression and individuals at high risk for depression. *Journal of the American Medical Association Psychiatry, 70*, 698–708. doi:10.1001/jamapsychiatry.2013.1189

Zimbardo, P. G., & Boyd, J. N. (2015). Putting time in perspective: A valid, reliable individual-differences metric. In *Time Perspective Theory; Review, Research and Application* (pp. 17-55). Springer International Publishing. Retrieved from: http://link.springer.com/chapter/10.1007/978-3-319-07368-2_2