Effects of Acute Psychosocial Stress on Declarative Memory for Neutral Material: A meta-analytic review

Bonny Astor (ASTBON001)
Nicole De Villiers (DVLNIC008)

ACSENT Laboratory
Department of Psychology
University of Cape Town

Supervisors: Kevin Thomas and Robyn Human
Word Count:
Abstract: [220]
Main Body: [9996]

Abstract

A vast body of research describes the effects of acute psychosocial stress on declarative memory for neutral, non-emotive verbal material. However, findings are equivocal as to whether these effects are enhancing, impairing, or non-existent. We used meta-analytic procedures to investigate potential moderating variables that might explain some of this inconsistency. We reviewed the data from 19 studies, drawn from 16 published articles, that experimentally investigated the effects of acute psychosocial stress on the immediate recall, delayed recall, and recognition of neutral verbal material. Though none of the investigated moderators (stage of memory process at which stress was administered; sex of sample; time of day at which the study was run), behaved in a manner consistent with our hypotheses, we observed an important (and unexpected) result regarding the operationalization of declarative memory. Specifically, the comparison between parallel analyses revealed that a unitary operationalization of declarative memory is likely to be more valid than a multi-component alternative. Stated otherwise, the meta-analysis suggested that when declarative memory operationalized by one outcome measure (viz., delayed recall) only is less confounded by within-study variability than when operationalized by multiple measures (viz., immediate recall, delayed recall, and recognition). Ongoing research into the effects of acute psychosocial stress on declarative memory is likely to be more effective, and more valid, if it takes this finding into account.

Effects of Acute Psychosocial Stress on Declarative Memory for Neutral Material: A meta-analytic review

Over the past half century, a vast body of research studying how stress influences cognitive performance has developed (McEwen & Sapolsky, 1995; Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012; Staal, 2004). Some of this research has demonstrated that chronic exposure to environmental stressors may have long-term consequences. For instance, chronic stress can result in enduring, dysregulated levels of the stress hormone cortisol, which is implicated in memory functioning. Such long-term effects are evident in major neurological and psychiatric disorders such as posttraumatic stress disorder and dementia of the Alzheimer's type (Alzheimer's Association, 2013; American Psychiatric Association [APA], 2013; Elzinga & Bremner, 2002).

Acute stress also has both short- and long-term implications for memory (Alderson & Novack, 2002). In fact, a major focus of investigation within the field of stress research is the way in which specific neurological processes, triggered by acute experiences of stress, affect memory at different stages of the cognitive process. These effects are of particular interest because there are many situations in which acute psychosocial stress is inherent and memory performance is critical (e.g., job interviews). However, though this area of the literature is notable for its depth, the findings are inconclusive and controversial (Het, Ramlow, & Wolf, 2005; Luethi, Meier, & Sandi, 2009; Sauro, Jorgensen, & Pedlow, 2003; Wolf, 2009). Hence, there is good rationale for continued investigation into both why and how memory is affected by exposure to acute psychosocial stress.

Conceptualizing Stress

'Stress' refers to a range of concepts including: (a) the challenging stimulus (the stressor), (b) the associated behavioural response, and (c) the physiological response (Kemeny, 2003). Stressors are defined as circumstances that can affect physiological and psychological wellbeing. Physiological responses to stressors are evolutionarily adaptive, allostatic mechanisms that ready the body to respond to threat (Joëls & Baram, 2009; Vingerhoets & Perski, 1999).

One specific physiological response to a stressor involves activation of the hypothalamic-pituitary-adrenal (HPA) axis. The HPA axis is a vital regulatory system connecting the central nervous system and the endocrine system. Hence, it not only allows the human organism to adapt under emotionally and physically stressful conditions, but also

regulates crucial functions, such as growth and reproduction, under basal conditions (Kirschbaum, Wüst, & Hellhammer, 1992; Kudielka & Kirschbaum, 2005).

HPA-axis activation in response to a stressor consists of three phases: 1) The *pre-stressor activity* phase (basal activity), 2) the *stress-reactivity* phase, which results in increased cortisol levels following a stressful situation, and 3) the *stress-recovery* phase following the offset of a stressor, during which cortisol levels return to baseline (Burke, Davis, Otte, & Mohr, 2005). Across these phases, the following physiological events occur: Corticotropin-releasing hormone is secreted from the hypothalamus, leading to the release of adrenocorticotropic hormone (ACTH) from the pituitary gland. ACTH then prompts the release of glucocorticoids (cortisol in humans) from the adrenal cortex (Alderson & Novack, 2002; Kudielka, Hellhammer, & Wüst, 2009; Kudielka & Kirschbaum, 2005; McEwen & Sapolsky, 1995). Once released, cortisol travels through the blood-brain barrier where it binds to receptors in a number of areas.

Stress and Memory

Glucocorticoid secretion in response to stress has consequences for particular brain structures with particularly high concentrations of glucocorticoid receptors, and ultimately results in modified memory function. Both chronic and acute adrenocortical activation appear to affect memory performance (de Quervain, Roozendaal, Nitsch, McGaugh, & Hock; 2000; Roozendaal, 2002). In humans, cortisol acts on two particular brain regions: the hippocampus and the prefrontal cortex (Alderson & Novack, 2002; Kemeny, 2003; Wolf, 2009). These areas are critical for declarative memory (Eichenbaum, 2003; Squire, 1992) and working memory (Oei, Everaerd, Elzinga, Van Well, & Bermond, 2006; Wolf, 2003), respectively. The impairing effects of stress on working memory are well documented and robust, but research into the effects of stress on declarative memory is much less conclusive, and therefore requires further investigation (Elzinga & Roelofs, 2005; Kirschbaum, Wolf, May, Wippich, & Hellhammer, 1996; Oei, et al., 2006; Roozendaal, 2002; Wolf, 2009).

Stress and declarative memory. Declarative memory is the explicit processing of specific information that is easily remembered (i.e., facts and events; Squire, 1992). The hippocampus plays an important role in declarative memory functioning. It also contains high concentrations of glucocorticoid receptors (McEwen & Sapolsky, 1995). Therefore, dysfunction of the hippocampus resulting from elevated glucocorticoid levels (such as occurs under stressful conditions) has implications for learning and recall of declarative material (de Quervain et al., 2000; McEwen & Sapolsky, 1995; Oei et al., 2006).

Numerous empirical studies, using either direct glucocorticoid administration or laboratory-based acute psychosocial stressors (e.g., the Trier Social Stress Test [TSST]; Kirschbaum, Pirke, & Hellhammer, 1993) to elevate cortisol levels, suggest that stress-induced cortisol secretion can affect declarative memory in variable ways (i.e., it can have enhancing, impairing, or no effects; de Quervain et al., 2000; Kirschbaum et al., 1996; Smeets, Jelicic & Merckelbach, 2006a). Several influential studies suggest that these contradictory effects might result from the fact that stress influences different stages of the memory process in different ways (Beckner, Tucker, Delville, & Mohr, 2006; de Quervain et al., 2000; Lupien & Schramek, 2006; Roozendaal et al., 2002).

Effects of stress at different stages of memory processing. Researchers conceptualize three main stages of memory processing: (1) encoding (learning information), (2) consolidation (storing information), and (3) retrieval (remembering information; de Quervain, Aerni, Schelling, & Roozendaal, 2009; Schwabe et al., 2012; Tulving, 1995). Roozendaal (2002) suggested that acute glucocorticoid administration and acute stress influence consolidation and retrieval in contrasting ways. By his theoretical model, increased levels of cortisol enhance memory consolidation and the storage of novel information, but impair memory retrieval. It is thought that these contrasting outcomes are due to different effects of stress-induced cortisol elevation on brain structures implicated at different stages of the memory process (see also Smeets, Giesbrecht, Jelicic, & Merckelbach, 2007).

Many recent studies have tested the predictions made by Roozendaal's (2002) model (i.e., they have investigated the variable effects of stress on declarative memory by isolating the specific stages of the memory process at which stress is applied). In these experimental paradigms, a stimulus set is presented and memory for that set is subsequently tested. Administration of the stressor may occur at any of the following stages of the experimental protocol: just before stimulus presentation (i.e., at encoding), during the interval between stimulus presentation and memory testing (i.e., during consolidation), or immediately prior to memory testing (i.e., at retrieval). However, although many investigations of effects at each stage have been conducted, there is no consensus in the literature.

Stress at encoding. Empirical studies suggest that when the psychosocial stressor is administered at the encoding stage of the memory process, declarative memory performance is either unaffected (Domes, Heinrichs, Reichwald, & Hautzinger, 2002; Domes, Heinrichs, Rimmele, Reichwald, & Hautzinger, 2004; Kuhlmann, Piel, & Wolf, 2005; Luethi et al., 2009; Lupien, Gillin, & Hauger, 1999; Nater et al., 2007; Smeets et al., 2006a; Wolf et al., 2001) or impaired (Elzinga, Bakker, & Bremner, 2005; Jelicic, Geraerts, Merckelbach, &

Guerrieri, 2004; Smeets et al., 2006a, Smeets, Jelicic, Merckelbach, 2006b; Smeets et al., 2007; Stawski, Sliwinski, & Smyth, 2009).

Stress at consolidation. Empirical studies suggest that when the psychosocial stressor is administered at the consolidation stage, declarative memory performance is either impaired (Stawski et al., 2009), unaffected (de Quervain et al., 2000; Elzinga et al., 2005), or enhanced (Smeets, Sjistermans, Gisjen, Peters, Jelicic, & Merckelbach, 2008).

Stress at retrieval. Most empirical studies suggest that when the psychosocial stressor is administered at the retrieval stage, declarative memory performance is impaired (Boehringer, Schwabe, & Schachinger, 2010; de Quervain et al., 2000; de Quervain et al., 2003; Domes et al., 2004; Kuhlmann et al., 2005). However, a small group of studies report no effect of stress on declarative memory at retrieval (Domes et al., 2004; Kuhlmann et al., 2005; Schoofs & Wolf, 2009).

Clearly, then, there are no consistent stage-dependent effects of stress on memory performance. In fact, findings from the studies reviewed above suggest that at each stage of the memory process, the effects of stress-induced cortisol elevations may impair, enhance, or have no effect on memory performance. It is important to make sense of this apparent conundrum in order to truly understand the effects of acute stress on declarative memory. One way of doing this is by examining variables that moderate the relationship between stress-induced cortisol elevation and memory performance.

Variables Moderating the Relationship between Stress and Memory

A distinct line of research has attempted to experimentally isolate variables that may moderate the relationship between stress and memory performance. Researchers have considered both participant- and context-related variables (de Quervain et al., 2000; Roozendaal, 2002; Smeets et al., 2006a).

Participant-related variables. Participant-related variables, such as sex, age, individual cortisol response, and personal habits (e.g., smoking, caffeine intake, exercise, etc.) have all been cited as potential moderators of the effect of stress on memory. However, most of these variables are consistently controlled for in most published studies, and are therefore unlikely to contribute significantly to between-study variance. Sex is an exception, however, because even though it might be well controlled for, its effect on memory is still unpredictable (Künzel et al., 2003; Nicholson, 1989; Sauro et al., 2003).

Sex of sample. Studies of sex differences in adrenocortical activity have revealed important differences between men and women. Specifically, it appears that men have higher

baseline cortisol levels than women (Kirschbaum, Kudielka, Gaab, Schomer, & Hellhammer, 1999; Kirschbaum et al., 1992; Schöneshöfer & Wagner, 1977). Additionally, laboratory-based psychosocial stressors tend to elicit higher cortisol responses in men (almost twice as high) than in women (Kirschbaum et al., 1992, 1993; Kudielka et al., 2009; Schoofs & Wolf, 2009).

Furthermore, cortisol response is much more variable in women than in men because the availability of free cortisol varies with phase of the menstrual cycle (Kirschbaum et al., 1992; Kirschbaum, Pirke, & Hellhammer, 1995). Consistent with this variability is the finding that menstrual cycle phase strongly influences memory performance under conditions of psychosocial stress. Men and women have comparable cortisol responses when women are in the luteal phase of the menstrual cycle (Espin et al., 2013; Kirschbaum et al., 1992; Kudielka et al., 2009). However, women in the follicular phase of the menstrual cycle and women taking oral contraceptives show, relative to controls, significantly reduced cortisol responses to acute psychosocial stressors (Kirschbaum, Pirke et al., 1995). Therefore, the interaction of sex, menstrual cycle phase, and oral contraceptive use affects stress responsiveness in women (Espin et al., 2013; Kirschbaum et al., 1999).

Studies that do not take these sex differences into account may be vulnerable to methodological confounds and flawed interpretation of data.

Context-related variables. Features of the experimental process such as timeline of experimental protocol, non-discrete stages of the memory process, measures of memory, stimuli valence, time of day, and means of stress elicitation, may also influence the effects of stress on declarative memory.

Experimental protocol timeline. On average, cortisol responses peak 20-40 minutes after administration of an acute stressor, but only revert to baseline levels 60-90 minutes after the stressor ends (Kudielka & Kirschbaum, 2005). Unless there is a sufficient delay between the learning of stimulus material and the recall of that material, participants may have elevated cortisol levels during more than one stage of the memory process. Whether or not this factor is taken into account may alter the interpretation of findings (Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006). Furthermore, whether or not the experimental protocol is completed in one session on one day, or takes place over multiple sessions at least 24 hours apart, may influence memory performance. Such influence arises particularly because healthy sleep is important for memory consolidation (Elzinga et al., 2005; Marshall & Born, 2007).

Non-discrete stages of the memory process. Because memory is a continuous and interactive mental phenomenon, the stages of the memory process described above

(encoding, consolidation, and retrieval) are artificial delineations (Walker, Brakefield, Hobson, & Stickgold, 2003). There are no known robust timeframes for these stages (e.g., when encoding ends and consolidation begins). In fact, some evidence suggests that stages (e.g., consolidation and retrieval) may occur simultaneously (Roozendaal, 2002; Smeets et al., 2007). Different researchers experimentally conceptualize stages of the memory process in different ways, contributing to further to inconsistency in protocols. This is likely to contribute to inconsistency in results.

Measures of memory. Several measures of verbal declarative memory (e.g., free recall, cued recall, or recognition) can be used in stress protocols. Free recall tasks require participants to recall learned information (e.g., word lists or stories) without the aid of prompts, whereas in cued recall tasks participants are presented with a list of items and are then prompted with target word stems during recall (Elzinga et al., 2005; Tulving & Pearlstone, 1966). Recognition tests simply require participants to indicate whether a word has been presented before or not, and therefore require less effortful processing (Haberlandt, 1999; Lupien et al., 1999). Each of the above tasks may be administered either immediately following learning, or after a filled delay of some particular length (Het et al., 2005; Sauro et al., 2003).

Valence of test stimuli. Memory performance under stress is also largely dependent on the valence of the stimuli being studied (i.e., neutral versus emotional [positive or negative]; de Quervain et al., 2000; Smeets et al., 2006b). Emotionally arousing material tends to be less impaired by psychosocial stress than neutral material, and in some instances is even enhanced. These effects may be explained by the fact that an important additional memory-modulating brain structure (the amygdala) is activated during encoding of emotional material (Cahill, 2003).

Time of day. The circadian cortisol rhythm is at its peak in the morning, when there is high HPA-axis activity. Low HPA-axis activity, on the other hand, typically occurs during the evening. Kuhlmann and colleagues (2005) reported negative effects on cortisol levels when a psychosocial stressor was administered during the morning. Furthermore, according to Lupien et al. (2002), glucocorticoid treatment has impairing effects on memory performance when administered in the morning, but enhancing effects when administered in the afternoon. However, even in highly controlled studies, individual differences in cortisol response are still reported. These differences may arise due to individual variation in circadian cortisol cycle: Some studies suggest that up to 17% of individuals exhibit no such cycle (Kirschbaum, Prussner, et al., 1995; Smyth et al., 1997).

Means of stress elicitation. Direct administration of cortisol is a highly effective pharmacological treatment with little variation in its effects on cortisol elevation. Psychosocial stressors, however, generate more variable stress responses than direct administration of cortisol. This increased variability exists at least partially due to the influence of participants' past experiences. For example, practice effects (e.g., experience with public speaking) are likely to attenuate the degree to which psychosocial stressors elicit cortisol responses (Dickerson & Kemeny, 2004; Het et al., 2005; Kirschbaum et al., 1996; Pickering & Gerin, 1990; Sauro et al., 2003). To reduce variability, many studies use the TSST (Kirschbaum et al., 1993). This standardized psychosocial stress elicitation paradigm reliably induces moderate stress. It is comprised of three components (motivated performance, social evaluation, and uncontrollability), each of which is essential to its ability to induce stress (Dickerson & Kemeny, 2004; Mason, 1968).

Rationale and Specific Aims

The literature in this area is extremely inconsistent with regards to the effects of psychosocial stress on declarative memory. To determine whether factors reviewed above (i.e., sex of sample, timeline of experimental protocol, measures of memory, and time of day) are important moderators of the effects of stress on memory, and therefore potential methodological confounds if not taken into account, the results of studies in this area must be analyzed thoroughly and systematically. We therefore conducted such an analysis using meta-analytic techniques.

Meta-analysis is a quantitative methodological approach for integrating multiple and potentially contradictory results within a particular area of study (Glass, 1976; Glass, McGaw, & Smith, 1981; Lipsey & Wilson, 2011; Rosenthal & Rosnow, 2008). In this method, standardized measures of effect (effect sizes) are generated for each of the included studies, enabling them to be compared directly. As the literature in this area is especially complex and findings are extremely inconsistent, a meta-analytic approach is valuable.

To reduce some complexity in this analysis, we specifically investigated the effects of acute psychosocial stress, as elicited by the TSST (or closely comparable procedures), on declarative memory performance. To further reduce discrepancy, we used verbal declarative memory for neutral material as the outcome of interest. We did so because both emotional and visual material implicate additional memory-modulatory brain structures and may confound the analysis (Cahill, 2003).

Based on this rationale and the literature reviewed above, the specific objectives of the current meta-analysis were as follows. The first aim was to quantitatively review the overall effect of psychosocial stress on verbal declarative memory for neutral material. The second aim was to determine whether there was variability in this effect that could be explained by moderator variables (i.e., sex of sample, timeline of experimental protocol, measures of memory, and time of day). The third aim was to investigate the amount of variance explained by these moderators, as opposed to by within-study confounds.

Method

Sample Search

We used a multi-staged, iterative procedure to obtain the final sample of studies that were included in the meta-analysis. This approach follows PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher, Liberati, Tetzlaff, & Altman, 2009) guidelines, and is illustrated in Figure 1.

The initial pool of primary studies was acquired by using online research databases (Medline and PsycINFO) using the Boolean phrase: *(free recall OR cued recall OR recognition OR memory) AND (stress OR cortisol OR TSST OR Trier Social Stress Test)*. We excluded the following keywords so as to refine search results: *post-traumatic stress, drug*, drug-related, autobiograph*, prenatal, maternal, depress*, dementia, psychosis, schizophren*, alzheimer*, (syndrome OR disorder)*. This search yielded 584 papers. Two researchers (B.A. and N.D.V) scrutinized this sample independently, and excluded studies that did not meet the selection criteria described in the following section.

Selection criteria for the sample. Papers included in the final sample of this meta-analysis were required to meet following criteria: (a) peer-reviewed journal article; (b) published in English; (c) purely psychosocial stressor (TSST or closely comparable); (d) published between 1993 and January 2014 (because the original account of the TSST was published in 1993); (e) quantitative methodology; (f) population of healthy adult humans, within the age range of 18-55 years; (g) use of an appropriate control group; (h) verbal encoding of information; (i) measure of verbal declarative memory for neutral words; (j) consistent laboratory-based context for learning and memory testing; and (k) measure of cortisol as a check on the stress manipulation[1].

These inclusion criteria were based on the following rationale. Cortisol reactivity is, at least in part, age-dependent, and therefore we excluded studies with samples consisting exclusively of participants over the age of 50 (Kudielka, Schmidt-Reinwald, Hellhammer, & Kirschbaum, 1999). Some psychiatric disorders (e.g., posttraumatic stress disorder) are implicated in cortisol dysregulation (Sauro et al., 2003), and therefore we included only studies with non-clinical samples. In studies featuring within-subjects designs, we included only those with parallel versions of each declarative memory test, in order to compensate for

[1] All included studies had to have a stress manipulation check: that is, one or several measures of cortisol as a means of showing that the stress manipulation did, in fact, lead to elevated cortisol levels and thereby resulted in modified memory functioning (de Quervain et al., 2000).

possible test-retest effects. We excluded studies that did not feature a control group, or that featured an inappropriate control group (e.g., a control group in which stress was administered less than 90 minutes before encoding or retrieval, which meant that stress was potentially affecting declarative memory). Finally, because we were only interested in hippocampal-dependent forms of memory, we excluded studies that tested encoding or retrieval of visual images (e.g., photographs or faces) and/or emotional material (positive or negative). The total number of studies that met these criteria was 50.

Study elimination. Four researchers (B.A., N.D.V., R.H., and K.T.) assessed the remaining 50 articles in detail. The decision to include or exclude a study from the final sample was made by consensus in line with the inclusion criteria. Eventually, 11 papers were retained. To ensure that all relevant articles were included in the final sample, the same four researchers scrutinized the reference lists of included studies independently, and determined by consensus that 5 additional studies were eligible for inclusion. We sub-divided one of the papers in the final sample (Smeets et al., 2006a) because it consisted of two separate independent and eligible studies. Hence, at this point the total sample of included studies was 17.

Splitting papers. Most papers in our sample investigated multiple memory measures and/or multiple treatment groups (n = 14). Multiple stress treatment groups, where stress was administered at different stages of the memory process (Domes et al., 2004; Smeets et al., 2009; Stawski et al., 2009; see Appendix A), were divided into separate studies (n = 6) and treated as such in the analysis, making the sample size 20. Furthermore, within some papers (Domes et al., 2004; Elzinga et al., 2005; Espin, 2013; Luethi, 2009; Smeets et al., 2006a, 2006b; Smeets et al., 2009; Stawski et al., 2009) multiple types of memory measure were tested; a total of 28 separate effect sizes were calculated for different types of memory measures within these studies. However, as representing a study more than once in a meta-analysis violates independence of samples, study components were weighted. When groups (i.e., stress or control) within studies were split, their sample sizes were reduced so that, in sum, they equaled the original sample size (Higgins & Green, 2008; Lipsey & Wilson, 2001; Rosenthal & Rosnow, 2008).

Coding

Three independent raters (B.A., N.D.V., and R.H.) coded each study in the final sample. They recorded the following information: (a) year of publication; (b) total number of participants; (c) experimental design (within-subjects or between-subjects); (d) total number of groups; (e) number of participants within each group; (f) sex of sample (male, female, or

mixed); (g) age (mean, standard deviation, and range); (h) stage psychosocial stressor was administered (encoding, consolidation, or retrieval); (i) number and type of memory measures (immediate free recall, delayed free recall, delayed cued recall, and/or recognition); (j) length of time between learning and recall; (k) duration of experiment (< 24 hours or > 24 hours); and (l) time of day at which the study was conducted (i.e., morning or afternoon). Differences between raters were resolved through discussion.

Calculation of Effect Sizes

The studies included in this meta-analysis compare memory performance across treatment groups (stress versus control) and have variable sample sizes. Therefore, the type of effect size that must be calculated for each study is a weighted standardized mean difference: Cohen's *d.* This statistic is calculated by dividing the difference between the mean of the control (non-stress) group and the experimental (stress) group, by the combined standard deviation of both groups (Hedges, 1982). To correct for small-sample bias, we applied the Hedges and Olkin (1985) correction (Hedges' *g*) to this effect estimate.

Studies often use one or multiple word lists in order to test memory performance. These word lists can vary for stress groups: for example, one group could be tested for word lists related to recall for personality-related words, and the other, recall for memory-related words. Some studies reported separate statistics for memory performance within stress and control groups (e.g., separate data for recall of personality words and memory words; Smeets et al., 2007). In such cases, means were averaged and standard deviations pooled to provide overall descriptive statistics for stress and control groups; this calculation was performed for each study independently by two researchers (B.A. and N.D.V.) to reduce the risk of human error.

Four studies (Domes et al., 2004; Smeets et al., 2006a; Smeets et al., 2009; Stawski et al., 2009) included more than one stress group (e.g., stress at encoding, stress at consolidation, stress at retrieval). For those studies, we calculated separate *d*'s for each stress group.

Nine studies had more than one memory measure (e.g., immediate recall, delayed free recall, delayed cued recall, and recognition). For those studies, we calculated separate *d*'s for each memory measure (see Appendix A for splitting rationale). To avoid violating assumptions of sample independence, we weighted the components of the studies that had more than one memory measure, but had the same participants, by altering their sample sizes in the manner described above (Lipsey & Wilson, 2001).

We used *Review Manager (RevMan)*, a freely available meta-analytic software program (http://tech.cochrane.org/revman/download), designed for conducting systematic Cochrane

Reviews, to derive effect sizes from each study using the reported descriptive statistics. In cases where the necessary information was not provided, we attempted to contact the authors for additional information. If we received no reply, as was the case for seven studies (Domes et al., 2002, 2004; Espin et al., 2013; Schoofs & Wolf, 2009; Smeets et al., 2007, 2009) we used *WebPlotDigitizer* (an online data extraction software; http://arohatgi.info/WebPlotDigitizer/app/) to estimate means and standard deviations from the figures provided in those studies. This extraction was done for each figure independently by two researchers (B.A. and N.D.V), and the values found by each researcher were averaged. In two instances (Domes et al., 2004; Schoofs & Wolf, 2009), authors provided missing information after *WebPlotDigitizer* had been used to extract data from graphs. In those cases, the provided and extracted data were highly correlated with each other, confirming the reliability of the software (r = 0.97). In one instance (Hidalgo et al., 2012), extraction was not possible because figure data were unclear; in another case (Nater et al., 2007), figures did not provide control-group data. Hence, these studies were excluded from further analysis. The total sample size of included studies at this stage was therefore 19.

Once calculated, effect sizes were interpreted by their size (small effect sizes as ≤ 0.20, moderate as ≤ 0.50, and large ≥ 0.80), their direction (negative effect sizes indicate impaired memory performance in the stress group as compared to the control group, whereas positive effect sizes indicate enhanced memory performance in the stress group as compared to the control group; Cohen, 1988), and their confidence intervals (whether they were greater than 0, less than 0, or spanned 0).

Analysis of Effect Sizes

We analyzed the study outcome variable (declarative memory performance) in two main ways corresponding to the hypotheses listed above. We used a random effects model because it is effective at detecting potential sources of heterogeneity (Field & Gillett, 2010; Rosenthal, 1995).

Investigating overall effect. To investigate the overall effect of stress on declarative memory, we calculated the average weighted effect size of all included studies. To determine whether this effect size was derived from a homogenous sample of studies, we performed a *Chi-square* test of heterogeneity (*Q*). Significant heterogeneity of effect sizes indicates that more variation than can be expected by chance alone is occurring between studies. In other words, if we found such significant heterogeneity, it would suggest that there are variables other than the primary variable of interest (stress) influencing declarative memory performance. The magnitude of heterogeneity for the primary outcomes (I^2) was interpreted

by following the Cochrane handbook guidelines: 0 – 40 (might be important), 30 – 60 (moderate), 50 – 90 (substantial), 75 – 100 (considerable) (Higgins & Green, 2008).

Investigating moderators. To investigate potential moderating variables, we categorized studies into sub-groups (e.g., stress at encoding, stress at consolidation, stress at retrieval) and compared the average effect sizes of these sub-groups with one another. Significant moderation effects manifest as heterogeneity of effect sizes between sub-groups. We conducted analyses of heterogeneity between categories of studies to address each of the hypotheses stated above.

There is likely to be more heterogeneity within a group of small studies due to small-sample effects (including publication bias), with more precise estimates provided by larger studies (all else being equal). However, RevMan weighs the contribution of each study to the overall effect size by the precision of the effect reported by the study, which is a function of sample size. Therefore, to counteract the under-powered chi-square statistic of *Q*, we used a more liberal *p*-value of .10 instead of the conventional .05 to analyze the results of homogeneity tests (Higgins & Green, 2008).

Results

Study Features

Twenty studies (published in 16 papers) investigating the effects of acute psychosocial stress on verbal declarative memory for neutral material met the inclusion criteria for this meta-analysis. The total sample size of all included studies amounted to 934 healthy adults aged between 18 and 68 years, with an overall mean age of 23.59 years (*SD* = ±3.23). We detail each of these variations below.

As Table 1 shows, all included studies investigated the effects of an experimental (stress) treatment versus a control treatment on participants' memory performance. However, the studies' experimental protocols varied in terms of: (a) design (i.e., between- versus within-subjects); (b) sex of sample (all-male, all-female, or mixed); (c) time of day administered (morning versus afternoon); (d) stage of the memory process at which stress was administered (encoding, consolidation, or retrieval); (e) materials used to test memory (word lists versus paragraphs), and (f) type of memory tested (immediate free recall, delayed free recall, recognition).

Design. One study administered both treatments (stress and control) to each participant at different stages of the experimental protocol in a crossover design (within-subjects experiment), whereas 19 studies randomly assigned each participant to either a stress group or a control group (between-subjects experiments).

Sex of sample. Three studies investigated all-female samples, 8 investigated all-male samples, and 9 investigated mixed-sex samples.

Time of day. Eight studies administered their protocol in the morning, 11 in the afternoon or evening, and 1 over the course of a day (stress protocols administered at several stages throughout the morning, afternoon, and evening; Luethi et al., 2009).

Stage of the memory process. Three studies administered stress at encoding and had an interval of > 90 minutes before retrieval. Ten studies administered stress at encoding and had an interval of < 90 minutes before retrieval. Two studies administered stress at consolidation (just after encoding). Five studies administered stress at retrieval.

Table 1

Study Features for Final Sample

	n			Age							
Study[a]	n_{Total}	n_{Stress}	$n_{Control}$	Range	*M*	(SD)	Sex[b]	Design[c]	Memory Measure[d, e]	Stressor Administration[f]	Time of day	*d*[d]
Boehringer et al. (2010)	51	33	18	18-31	24.57	0.61	M	B	DR	Retrieval	PM	d^{DR} = -0.32
Domes et al. (2002)	32	20	12	32-68	47.30	10.30	F	B	DR	Encoding (<90)	AM	d^{DR} = -0.24
Domes et al. (2004) S1	40	20	20	18-42	25.30	6.60	M	B	DR & Recog	Encoding (>90)	PM	d^{DR} = -0.09 d^{Recog} = 0.04
Domes et al. (2004) S2	40	20	20	18-42	25.30	6.60	M	B	DR & Recog	Retrieval	PM	d^{DR} = -0.13 d^{Recog} = 0.00
Elzinga et al. (2005)	16	16	16	-	21.40	2.10	F	W	IR & DR	Encoding (<90)	PM	d^{IR} = -0.22 d^{DR} = -0.19
Espin et al. (2013)	119	57	62	18-25	19.33	1.77	M+F	B	IR & DR (RAVLT)	Encoding (<90)	PM	d^{IR} = 0.20 d^{DR} = 0.43
Jelicic et al. (2004)	40	20	20	-	20.10	-	M+F	B	IR (AVLT)	Encoding (<90)	PM	d^{IR} = -0.36
Kuhlmann et al. (2005)	19	19	19	19-40	24.58	1.26	M	B	DR	Retrieval	AM	d^{DR} = 0.09
Luethi et al. (2009)	35	19	16	30-34	23.40	2.90	M	B	DR & Recog	Encoding (<90)	AM+PM	d^{DR} = -0.41 d^{Recog} = -0.51
Schoofs & Wolf (2009)	36	36	36	-	24.47	0.63	F	B	DR	Retrieval	AM	d^{DR} = 0.19

Smeets et al. (2007)	52	34	18	-	23.09	3.81	M+F	B	DR	Encoding (>90)	AM	d^{DR} = 0.25
Smeets et al. (2006a) S1	60	30	30	-	19.91	3.32	M+F	B	IR & Recog (DRM paradigm)	Encoding (<90)	PM	d^{IR} = -0.05 d^{Recog} = 0.21
Smeets et al. (2006a) S2	92	68	24	-	19.74	1.87	M+F	B	IR & Recog (DRM paradigm)	Encoding (<90)	PM	d^{IR} = -0.65 d^{Recog} = -0.30
Smeets et al. (2006b)	60	30	30	17-28	19.65	0.24	M+F	B	DR & Recog (30WVLT)	Encoding (<90)	PM	d^{DR} = -0.85 d^{Recog} = -0.91
Smeets et al. (2009) S1	32	16	16	18-39	20.70	3.30	M	B	DR	Consolidation	AM	d^{DR} = -0.36
Smeets et al. (2009) S2	32	16	16	18-39	20.70	3.30	M	B	IR & DR	Encoding (>90)	AM	d^{IR} = -0.27 d^{DR} = -0.33
Stawski et al. (2009) S1	50	25	25	18-24	18.94	1.02	M+F	B	DR	Consolidation	PM	d^{DR} = -0.18
Stawski et al. (2009) S2	50	25	25	18-24	18.94	1.02	M+F	B	IR	Encoding (<90)	PM	d^{IR} = -0.57
Tollenaar et al. (2008)[g]	-	-	-	-	-	-	-	-	-	-	-	-
Wolf et al. (2001)	58	22	36	-	24.25	0.83	M+F	B	DR	Encoding (<90)	AM	d^{DR} = -0.07

Note. [a]Where papers were split into multiple studies, S1 = the first identified study from the paper, S2 = the second identified study from the paper. [b]M = Male-only sample, F = Female-only sample, M+F = Mixed sample. [c]B = Between-subjects, W = Within-Subjects. [d]IFR = Immediate Recall, DR = Delayed Recall, Recog = Recognition. [e]Unless otherwise stated, all memory outcomes are measured using a word list. [f] (<90) = delay of less than 90 minutes between end of stressor and retrieval, (>90) = delay of more than 90 minutes between end of stressor and retrieval. [g]This study was excluded due to being the only study in the final sample investigating the effects of stress on cued recall.

Materials used to test memory. Three studies used the Rey-Auditory Verbal Learning Test (RAVLT; Taylor, 1959) or a modified version thereof, 1 study administered the Logical Memory subtest from the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987), 2 studies administered the Deese-Roediger-McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm, and the remaining 14 studies used non-standardized word lists. Many studies tested memory performance on both emotionally-valenced material and the neutral material of interest to this meta-analysis. We considered only the data relating to memory for neutral material.

Type of memory tested. Two studies tested both immediate and delayed free recall, 4 tested just immediate free recall, and 13 tested just delayed free recall. Recognition was also tested in 6 of the studies measuring free recall. If multiple repetitions of the same test were administered, we included only the first measure. Only one study (Tollenaar et al., 2008) tested cued recall and so was excluded at this stage, making the total sample size 19.

Primary Analysis

To estimate the overall effect of stress on declarative memory performance, we integrated all 28 calculated effect sizes. This integration resulted in a weighted average effect size of d = -0.14, with a 95% confidence interval of $-0.27 \leq d \leq -0.00$. This indicates a small impairing effect of stress on declarative memory. The Q_T value was 26.16 (p = .51), which indicates that there was statistically significant homogeneity in the total sample of studies. Furthermore, none (I^2 = 0%) of the total variability in this group of studies is explained by between-study heterogeneity. This suggests that the sample of studies come from the same population (Field, 2010; Rosenthal, 1995). However, a finding of statistically significant homogeneity does not preclude investigation of potential moderators (Rosenthal & Rosnow, 2008). Moreover, the number of papers included in this meta-analysis is small (k = 16), increasing the likelihood that between-study heterogeneity is going undetected (Type II error; Higgins & Green, 2011). Based on these considerations, we conducted further analyses to investigate our hypotheses (see Figure 2 for the forest plot of the primary analysis).

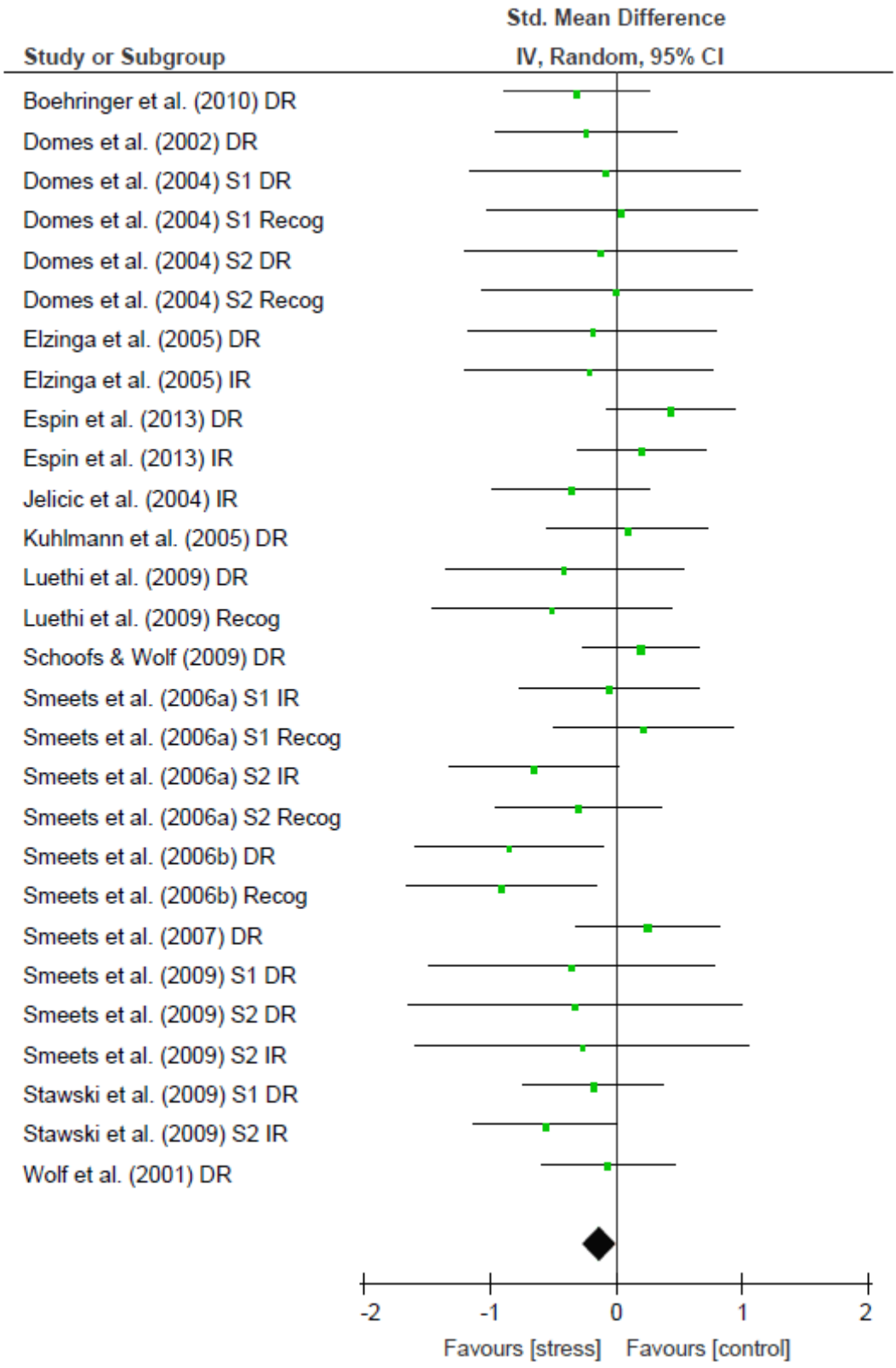

Figure 2. Forest plot showing the effect sizes and confidence intervals of studies included in the primary analysis.

Hypothesis testing. Consistent with our hypotheses, we conducted the following moderator analyses to investigate: (1) measure of memory (comparing: immediate, delayed, recognition), (2) stage of memory process at which stress was administered (comparing: encoding, consolidation, retrieval), (3) sex of sample (comparing: male, female, mixed), and (4) time of day (comparing: morning, afternoon). For each of these comparisons, we categorized each study into its appropriate group (if the relevant information was available), and then compared effect sizes and heterogeneity across categories. We excluded one study (Luethi et al., 2009) from the time of day comparison because its protocol stretched over both morning and afternoon, and it therefore did not fit into either category.

Table 2

Results for Primary Analysis

	Memory performance				
Category	*d's*	*d*	95% CI	Q_T	I^2
Measure of memory:					
Immediate recall	7	-0.25	-0.51, 0.01	5.97 (p = .43)	0%
Delayed recall	15	-0.05	-0.22, 0.13	12.50 (p = .57)	0%
Recognition	6	-0.27	-0.62, 0.08	5.32 (p = .38)	6%
Between category			$Q_B(2)$ = 2.33, p = .31		14 %
Stage of memory:					
Encoding	21	-0.14	-0.30, 0.02	19.35 (p = .50)	0%
Consolidation	2	-0.21	-0.72, 0.28	0.07 (p = .79)	0%
Retrieval	5	0.00	-0.29, 0.30	1.94 (p = .75)	0%
Between category			$Q_B(2)$ = 0.89, p = .64		0%
Sex of sample:					
Male	11	-0.20	-0.47, 0.07	2.18 (p = .99)	0%
Female	4	0.00	-1.34, 0.34	1.41 (p = .70)	0%
Mixed	13	-0.18	-0.51, 0.36	21.87 (p = .04)	45%
Between category			$Q_B(2)$ = 0.95, p = .62		0%
Time of day:					
Morning	8	0.03	-0.21, 0.27	2.65 (p = .92)	0%
Afternoon	18	-0.21	-0.39, -0.02	20.17 (p = .27)	16%
Between category			$Q_B(1)$ = 2.39, p = .12		58.1%

Note. d's = number of effect sizes per category. *d* = Cohen's *d* effect size. CI = Confidence Interval. Q_T = test of within-category homogeneity. I^2 = magnitude of variability that can be explained by between-category variance. Q_B = test of between-category homogeneity.

Across all comparisons, the only group of studies for which a statistically significant difference was observed between the experimental and control groups was for studies conducted during the afternoon. Across all comparisons, the only sub-category that was statistically significantly heterogeneous was the mixed-sex sample of studies. No between sub-category heterogeneity was observed (see Table 2).

Within the *measure of memory* comparison, lack of heterogeneity between studies suggests that immediate recall, delayed recall, and recognition can be considered together as a valid, unitary, operationalization of declarative memory. However, this result was particularly surprising for the following reasons: (a) as mentioned before, our sample size is small, thereby increasing the chances of Type II error and the likelihood that heterogeneity of effects of stress on different memory measures is going undetected; and (b) previous studies and theory suggest that the effects of stress on these measures *are* heterogeneous (Het et al., 2005; Stawski et al., 2009). Therefore, we decided to conduct a sensitivity analysis for studies using delayed recall as an outcome measure (Higgins & Green, 2011). This analysis was not repeated for immediate recall or recognition, due to their underrepresentation in the

total sample (immediate recall, $k = 7$, and recognition, $k = 6$) and their uneven distribution across categories (e.g., there were no studies that used immediate recall measures and applied stress at consolidation or encoding).

Sensitivity Analysis for Delayed Recall

We re-calculated the total average effect size by integrating only those effect sizes from studies using delayed recall as an outcome measure. This calculation resulted in a weighted average effect size of $d = -0.11$ ($-0.30 \leq d \leq 0.08$) indicating no evidence of an effect of stress on delayed recall. The Q_T value was 22.11 ($p = .08$), which indicates that there was significant heterogeneity in this sample of studies. Furthermore, as indicated by the I^2 value, 37% of the total variability in this group of studies is explained by between-study heterogeneity (Higgins & Green, 2011; see Figure 3 for the forest plot of the primary analysis).

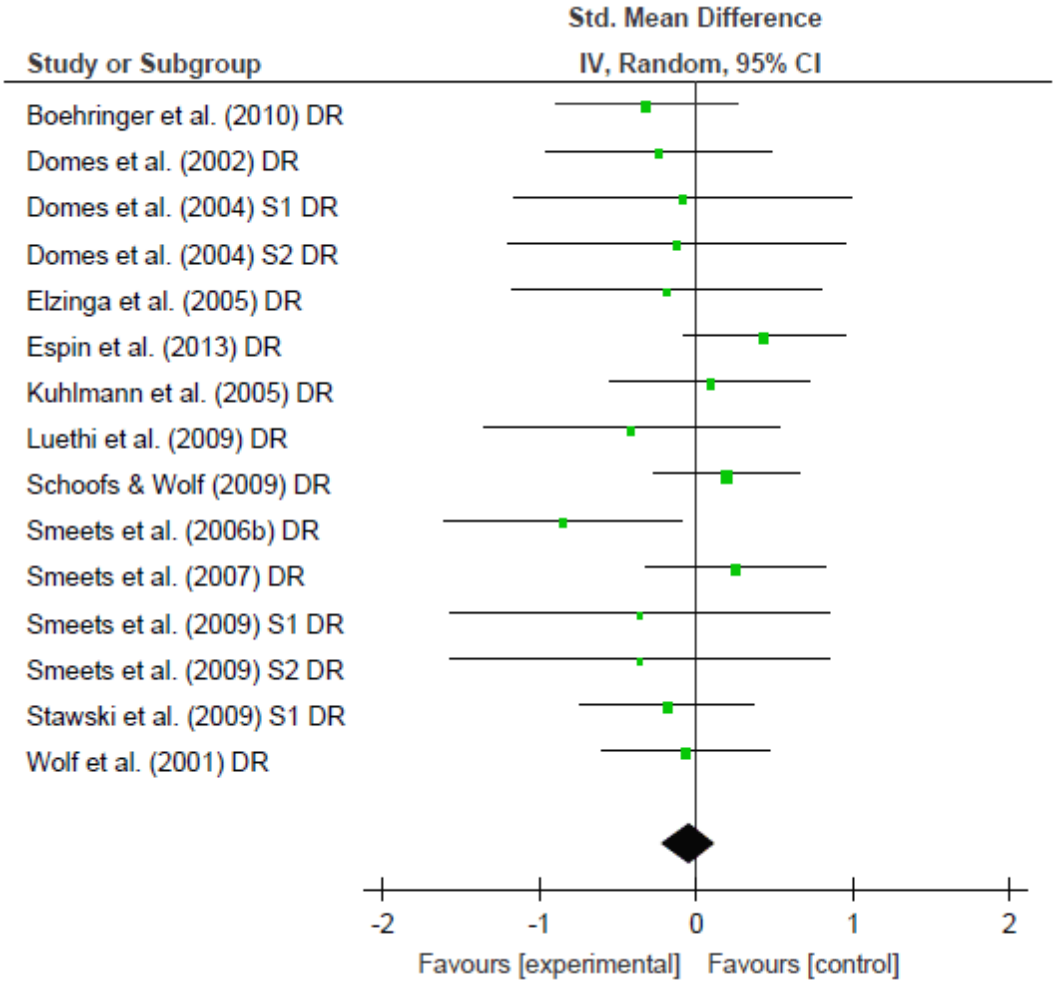

Figure 3. Forest plot showing the effect sizes and confidence intervals of studies included in the sensitivity analysis.

Hypothesis testing. The same comparisons to investigate hypotheses as run in the primary analysis were re-run on only delayed recall studies. For all of these comparisons, the confidence intervals of effect sizes indicated that there was no significant difference between stress and control groups (see Table 3).

Table 3
Results for Sensitivity Analysis

	Memory performance				
Category	*d's*	*d*	95% CI	Q_T	I^2
Stage of memory:					
Encoding	9	-0.10	-0.38, 0.18	14.57 (p = .07)	45%
Consolidation	2	-0.21	-0.57, 0.15	0.12 (p = .73)	0%
Retrieval	4	0.00	-0.30, 0.28	2.00 (p = .57)	0%
Between category			$Q_B(2) = 0.79$, p = .68		0%
Sex of sample:					
Male	7	-0.22	-0.49, 0.05	1.78 (p = .94)	0%
Female	3	0.00	-0.34, 0.34	1.38 (p = .50)	0%
Mixed	5	-0.07	-0.51, 0.36	17.31 (p = .002)	77%
Between category			$Q_B(2) = 1.14$, p = .56		0%
Time of day:					
Morning	8	0.03	-0.21, 0.26	2.88 (p = .82)	0%
Afternoon	7	-0.18	-0.52, 0.17	17.12 (p = .009)	65%
Between category			$Q_B(1) = 0.86$, p = .35		0%

Note. d's = number of effect sizes per category. *d* = Cohen's *d* effect size. CI = Confidence Interval. Q_T = test of within-category homogeneity. I^2 = magnitude of variability that can be explained by between-category variance. Q_B = test of between-category homogeneity.

Stage of memory. The results of this categorical integration indicate that stress had no effect on memory performance at any of the stages of the memory process. Studies in which stress was applied at encoding were significantly heterogeneous. To investigate this heterogeneity, we further stratified this category of studies by duration of delay between encoding and retrieval (> 90 minutes versus < 90 minutes) because, as hypothesized, this was expected to moderate the effects of stress on memory within the sub-category of encoding. For this comparison, the category of studies with delay of > 90 minutes was homogeneous (Q_T = 1.83, p =.40) and the stress manipulation had no effect (d = -0.03, 95% CI = $-0.39 \leq d \leq 0.33$) on delayed recall. The category of studies with delay of < 90 minutes, on the other hand, was very heterogeneous (Q_T = 17.37, p =.004), but again the stress manipulation had no effect (d = -0.20, 95% CI = $-0.65 \leq d \leq 0.25$) on delayed recall.

Sex of sample. Both sub-categories of male studies and of female studies were very homogeneous, whereas the sub-category of mixed sex studies was very heterogeneous. The

stress/control manipulation had no effect on delayed recall performance in any of these groups.

Time of day. The stress/control manipulation had no effect on declarative memory performance in studies conducted in the morning, or on those conducted in the afternoon. The category of morning studies was homogeneous, whereas afternoon studies were very heterogeneous.

Discussion

The objectives of this meta-analysis were to (1) quantitatively review the overall effect of acute psychosocial stress on verbal declarative memory for neutral material, (2) determine whether there was variability in this effect that could be explained by particular moderator variables, and (3) investigate the amount of variance explained by such moderators, as opposed to within-study confounds.

Summary and Implications of Results

The most interesting finding of this meta-analysis lies in the difference between the primary and sensitivity analyses. These analyses differed in the way that they operationalized declarative memory: the primary analysis operationalized it as a combination of immediate recall, delayed recall, and recognition, whereas the sensitivity analysis operationalized it as delayed recall only.

Overall, the primary analysis suggests that acute psychosocial stress has an *impairing effect* on declarative memory for neutral material. This impairing effect is indicated by the fact that the confidence interval of the weighted average effect size of all 19 studies included in the sample encompassed only negative numbers (where negative effect sizes indicate that stress groups performed more poorly than control groups), $-0.27 \leq d \leq -0.001$. In this analysis, the homogeneity within the total sample of studies, in combination with the large confidence intervals associated with each study (shown in Table 1 and Figure 2), suggest there is a high degree of within-study variability. Such variability is likely due to the presence of confounds in the methods of included studies, rather than to the systematic influence of moderators (Moayyedi, 2004). To researchers in this field, this finding indicates that studies investigating the acute effects of psychosocial stress on memory performance are complex, featuring many different moderators of unknown influence, and are therefore likely to be

confounded and their results difficult to interpret. In short, our finding goes some way toward explaining why there is such inconsistency in the literature on acute psychosocial stress and memory performance. We suggest, on the basis of this primary analysis, that there is an urgent need for standardization of study protocols.

Results from the secondary analysis provide further impetus for such standardization. Our rationale for conducting the secondary (sensitivity) analysis was that we expected it would allow closer examination of true effects. In other words, we were seeking a way to peel away at least one confounding layer that produced the high within-study variability that characterized the sample of studies within the primary analysis. Peeling away that layer, however, did not reduce opacity and magnify the strength of our focus on true effects; rather, it revealed multiple more opaque layers and potential sources of confound that can, potentially, obscure the true effects.

In the sensitivity analysis, we simplified the operationalization of declarative memory performance by removing all studies that used immediate recall or recognition as an outcome variable, and by retaining only studies that used delayed recall as an outcome variable. The results of this analysis suggested that acute psychosocial stress has *no effect* on declarative memory (operationalized in this simplified fashion) for neutral material.

Taking the sensitivity analysis further, we noted that the sub-sample of studies included in that analysis was heterogeneous. This heterogeneity indicates that more variance than can be explained by within-study confounds alone is occurring, and that moderating effects of other variables are likely. However, the outcomes of subsequent moderator analyses revealed an alternative explanation for this heterogeneity. For each comparison we conducted, there were *within-category* differences in the heterogeneity/homogeneity of sub-categories of studies (e.g., studies conducted in the morning were homogeneous, whereas those conducted in the afternoon were heterogeneous), even though there was no *between-category* heterogeneity in any of these comparisons (i.e., stress was not moderated by time of day). The sensitivity analysis therefore suggests that although there are differences in the heterogeneity/homogeneity of sub-categories of studies, none of the potential moderators we investigated had moderating effects. Rather, this analysis makes a significant contribution to the field by identifying specific sub-categories (e.g., afternoon

studies) in which within-study variation is large, highlighting the need for researchers to pay particular attention to the experimental designs of these studies.

Accurate operationalization of declarative memory is important. Comparing the primary to the secondary analyses reveals a clear difference in terms of explanatory power when a unitary operationalization is used (as in the sensitivity analysis, which used delayed recall only) versus when a multi-component operationalization is used (as in the primary analysis, which used immediate recall, delayed recall, and recognition). In the sensitivity analysis, the total sample proved to be significantly heterogeneous, and clear effects within sub-categories were apparent. In the primary analysis, the total sample was significantly homogeneous, and effects within sub-categories were far less apparent. And it seems the reason why effects are clouded in the primary analysis, but clear in the sensitivity analysis, is due to how truly each operationalization represents declarative memory.

Perhaps the multi-component operationalization of declarative memory is problematic because its constituent components are supported by different brain structures. Neuropsychological literature regarding memory-modulatory brain structures suggests that recognition tasks activate prefrontal, parietal, and medial temporal regions (Yonelinas, Otten, Shaw, & Rugg, 2005).[2] Similarly, immediate recall tasks activate prefrontal cortical structures such as those usually implicated in working memory processing, as well as medial temporal structures such as those usually implicated in declarative memory processing (Oei et al., 2006; Wolf, 2003).[3] Delayed recall tasks, on the other hand, are largely hippocampal-dependent, and so, in some ways, can be considered the 'purest' measure of declarative memory (Alderson & Novack, 2002).

In light of this literature, it is probable that operationalizing declarative memory using delayed recall (as done in the studies that were included in the sensitivity analysis) will allow clearer interpretation of the effects of acute psychosocial stress on declarative memory performance. However, it is impossible to determine from this meta-analysis whether delayed recall is the *best* operationalization of declarative

[2]This is because, typically, recognition memory depends on neural regions that support both autonoetic and noetic awareness (Tulving, 1985; Wheeler, Stuss, & Tulving, 1997).

[3]This is because, in immediate recall paradigms, learning is tested immediately following encoding, which increases the likelihood that some learned information is recalled from working memory (Jelici et al., 2004; Stawski et al., 2009)

memory, because the other two potential unitary operationalizations (just immediate recall, or just recognition) could not be investigated due to their uneven distribution across subsequent sub-categories of comparison. Therefore, although identifying the best operationalization of declarative memory was not one of the aims of this paper, the effects we uncovered have important implications for the field.

For instance, and this harks back to our earlier call for standardization of study protocols within this field, researchers might adopt what seems to be the best practice in operationalizing their outcome variables and select delayed recall as the sole (or, at least, a distinct) performance measure. This would be one step toward improving interpretability of findings, thus removing an obstructive opaque layer and moving closer toward inspection of true effects.

Meta-analytic investigation of potential moderating variables. Based on this idea that a unitary operationalization of declarative memory should be used in studies in this field, we will discuss investigation of potential moderating variables in terms of the outcomes of the sensitivity analysis.

None of the potential moderators investigated were shown to have a statistically significant effect. This was indicated by the absence of significant *between-category* heterogeneity in all of the moderator analyses conducted. However, within each moderator analysis, unexpected heterogeneity was found *within* one of the sub-categories. Heterogeneity within a sub-category indicates that there are important differences between studies included in this category (Rosenthal & Rosnow, 2008). The instances of heterogeneous sub-categories are discussed with regard to the specific moderator analysis in which they occurred:

***Stage of memory process*.** As suggested by inconsistency in the literature, but in contrast with influential theory (Roozendaal, 2002), the stage of the memory process at which stress was administered did not significantly moderate the effects of stress on memory. In this comparison, the category of studies in which stress was administered at encoding was significantly heterogeneous. Various findings in the literature suggest that stress administered at the encoding stage of the memory process is likely to have a different effect on memory when there is a delay of less than 90 minutes between encoding and retrieval compared to when there is a delay of more than 90 minutes (Kirschbaum et al, 1993). To investigate whether this might explain some of the

heterogeneity within this category, we stratified studies based on the duration of delay between encoding and retrieval in their respective procedures.

Stratification of encoding. In this stratification, duration of delay was not found to have a moderating effect. However, the sub-category of studies in which there was a sufficient delay was homogeneous, whereas the category of studies with an insufficient delay was heterogeneous. Such differences in within-category variance may be explained by the following findings within the literature: With a sufficient delay, cortisol levels have time to return to baseline before retrieval, whereas with an insufficient delay they may still be elevated at retrieval, thereby potentially also affecting the retrieval stage of the memory process, which is likely to confound effects (Burke et al., 2005).

Sex of sample. Sex of sample was not found to moderate the effects of acute psychosocial stress on declarative memory performance. However, in this moderator analysis, the mixed sex category of studies was very heterogeneous, whereas both female and male categories were homogenous (Kirschbaum et al., 1992).

The homogeneity of the female category of studies was not in line with the literature reviewed in this paper. Findings within the field suggest that this sub-category should be heterogeneous, due to confounding variables associated with female stress-induced cortisol response (e.g., menstrual cycle phase; Espin et al., 2013; Kirschbaum et al., 1992, 1999; Kudielka et al., 2009). However as there were only 3 studies in the female category, this finding should be interpreted with caution and more studies of this kind (i.e., with controlled female samples) must be done before conclusions may be drawn.

Heterogeneity within the mixed sex category was unexpected. A close examination of studies in this category revealed potential confounds in their methods which might be responsible for within-category heterogeneity: 1) some of the mixed studies have more male participants than female participants (e.g., Wolf et al., 2001), whereas others have more female participants than male participants (e.g., Espin et al., 2013; see Table 1); 2) within mixed studies, males and females were not always evenly distributed across treatment groups (i.e., stress versus control; e.g., Smeets et al., 2006a). Such methodological variance within this category is likely to contribute to heterogeneity within it, as the significant effects related to sex are not controlled for (Kirschbaum et al., 1992; Kudielka et al., 2009). In future research it is recommended

that studies a) use standardized 'mixed sample designs' (i.e. where males and females are equally represented and equally distributed across treatment groups), or b) report separate data for male and female outcomes.

Time of day. The time-of day at which studies were conducted was not found to moderate the effects of acute psychosocial stress on declarative memory performance. However, in this moderator analysis, studies conducted in the morning were homogenous, whereas there was very significant heterogeneity amongst studies conducted in the afternoon. The within-category heterogeneity of studies conducted in the afternoon is unexpected and not explained by the existing literature (Maheu, Collicutt, Kornik, Moszkowski, & Lupien, 2005). It is therefore likely that within-category heterogeneity is due to confounds between these studies. Therefore, further specific investigation of studies conducted in the afternoon is necessary before the heterogeneity of this category may be interpreted. This could easily be done by experiments whereby identical protocols are conducted in the morning for both stress and control groups, and in the afternoon for an additional stress group and an additional control group. Such experiments would be a valuable addition to literature regarding the effects of acute stress on memory performance, because they would add clarity to the existing controversies regarding the cortisol cycle, and help to explain the heterogeneity within this sub-category of studies.

It is notable that many of the effects observed in the above moderator analyses, derived from the sensitivity analysis, are also suggested in the primary analysis, but to a far *lesser* degree. This corroborates the interpretation that important effects of stress on declarative memory indeed appear to be clouded when a multi-component operationalization of declarative memory is used. Within sub-category effects only become clear when a unitary-component operationalization is used.

Limitations of the Current Study

All meta-analyses are limited by several inherent problems. There are 4 key limitations that need to be considered in the context of our meta-analysis. The first three are directly related to the strict selection (inclusion and exclusion) criteria that we used to acquire our sample of papers. The fourth is related to the methods used to report on study results.

Publication bias. The decision to only include previously published research in this meta-analysis introduces a form of *publication bias* because of the likelihood that papers with significant results (e.g., a significant impairing effect of psychosocial stress on memory performance) will be over-represented in the sample of studies. This issue of overrepresentation arises from the fact that papers without statistically significant findings are less likely to be published than those with statistically significant results (Moayyedi, 2004).

However, although this meta-analysis only included previously published studies, (thereby excluding any *grey literature*), the fact that a number of the studies (*n* = 9; 52.63%) included in our sample reported results that were *not* statistically significant indicates that our sample is likely to be representative. Therefore, the risk of publication bias negatively affecting our findings is unlikely.

The 'trash-in, trash-out' problem. This second issue, which is inherent to all meta-analyses, refers to the fact that the results produced by a meta-analysis are only as good as the results entered into the meta-analysis. This may result in the potential exaggeration of results through methodological flaws. In cases of included 'trash', biases are perpetuated in a meta-analysis by combining a series of inadequately conducted studies, thereby magnifying their shortcomings (i.e., the fact that they did not control for moderator variables; (Moayyedi, 2004).

We attempted to counter the effects of this problem by using strict inclusion criteria. For example, we required all included studies to be published in peer-reviewed journals. This reduced the likelihood of methodological flaws making their way into our meta-analysis. As our study sample consists of studies which are therefore of a high standard means that our meta-analysis is more likely to produce sound results.

Small sample size. The third limitation of our meta-analysis is that our sample size is relatively small (*n* = 19). This increases the importance of exercising caution when interpreting our results (Glass, 1976).

In this respect, our stringent inclusion and exclusion criteria may have been, in part, responsible for this limitation. However, the total sample size of included studies is similar to that of previous meta-analyses in the acute stress research field. These studies have reported similar or a smaller sample sizes (e.g., Het et al., 2005).

The lack of methodologically homogenous studies within this field has limited the scope of this meta-analysis. This existence of this limitation enables us to reiterate

the need for more, standardized, studies investigating the effects of acute psychosocial stress on verbal declarative memory.

Inaccessibility of reported data. In meta-analyses, some studies might be excluded for reasons other than the stated eligibility criteria (i.e., due to inaccessibility of the data).

Meta-analysis requires the use of data for each eligible study. are accessible to the researcher. Unfortunately, the studies included in this meta-analysis were found to be lacking in both the type and amount of data that were included.

Two studies had to be excluded due to a lack of available information. One of these study's did not report data for all groups, and the other study's plots were unreadable. Although repeated attempts were made to contact the authors of these studies, we did not receive any responses to our requests for additional information.

It is therefore important to highlight this limitation for all future published protocols. All data need to be reported consistently and comprehensively in tables of in figures (if data are not reported in tables, they should be provided in appendices), in order to avoid the need to exclude studies from future meta-analytic investigations (and could therefore lead to a bias in the results).

Strengths of Meta-Analysis.

Despite all of the above-mentioned limitations, meta-analysis as a technique has many strengths. It imposes a discipline on the analysis of literature within a field and by doing so enables important effects across studies to become apparent. This enables more sophisticated interpretations of findings to be made than are possible in many other types of study (e.g., systematic reviews; Higgins & Green, 2008). The discoveries of the present meta-analysis demonstrate these strengths of meta-analysis. Findings reveal important effects previously undiscovered within this field and by doing so make an invaluable contribution to it.

APPENDIX A

Rationale for Splitting Papers

The following papers were split into further studies for the following two reasons: The studies either had more than one stress group (i.e., the administration of the stressor at separate stages of the memory process), or, based on the encoding and retrieval of two different word lists at different stages of the memory process.

Domes et al., (2004) had two stress groups: *pre-learning stress* and *pre-retrieval stress*. They were split into two studies (i.e., (i) Domes et al., (2004) pre-learning stress, and (ii) Domes et al., (2004) pre-retrieval stress) because the stressor in each group was administered at two different stages of the memory process (encoding and retrieval, respectively).

Elzinga et al., (2005) was a within-groups, crossover design. The experimental protocol had multiple declarative memory tests (immediate free recall, delayed free recall, cued recall, and recognition), several versions of each test (e.g., word lists 1, 2 and 3; and paragraphs 1, 2 and 3), and was conducted over two days. However, due to the complexity of the study's experimental design, only immediate and delayed recall of paragraph 3 on day 1 could be included as the memory outcome measure for the stress group. The memory outcome measure for the control group was immediate and delayed recall of paragraph 1, because it occurred before the stressor, and on the same day as retrieval of paragraph 3. None of the other groups had comparable control groups.

Smeets et al., (2009) also had two stress groups: *pre-stress learning* and *post-stress immediate learning*, as well as an appropriate control group- this group was exposed to the TSST, but only encoded the word list 2 hours after the TSST ended, thereby giving enough time for cortisol levels to have returned to baseline). This paper was therefore sub-divided into two separate studies: (i) Smeets et al., (2009) pre-stress learning group, who were exposed to the stressor at consolidation, and (ii) Smeets et al., (2009) post-stress immediate learning group, who were exposed to the stressor at encoding.

Stawski et al., (2009) had two word lists, and tested delayed recall and immediate recall of both word lists. However, word list 1 was encoded and immediately recalled prior to the stress manipulation, whereas delayed recall of list 1 was tested after the manipulation. This means that the stressor was in no way affecting memory

processing at immediate recall, and so only delayed recall for word list 1 was included. Word list 2, however, was encoded, immediately recalled, and then tested for delayed recall all after the stress manipulation. In this case, both immediate and delayed recall were taken into account for list 2.

Furthermore, separate *d*'s (effect sizes) were also calculated for studies that had more than one memory measure (e.g., immediate recall, delayed free recall, and recognition). For example, Espin et al., (2013) tested immediate free recall and delayed free recall, and so separate effect sizes were calculated for each of these memory measures.

APPENDIX A

Rationale for Splitting Papers

The following papers were split into further studies for the following two reasons: The studies either had more than one stress group (i.e., the administration of the stressor at separate stages of the memory process), or, based on the encoding and retrieval of two different word lists at different stages of the memory process.

Domes et al., (2004) had two stress groups: *pre-learning stress* and *pre-retrieval stress*. They were split into two studies (i.e., (i) Domes et al., (2004) pre-learning stress, and (ii) Domes et al., (2004) pre-retrieval stress) because the stressor in each group was administered at two different stages of the memory process (encoding and retrieval, respectively).

Elzinga et al., (2005) was a within-groups, crossover design. The experimental protocol had multiple declarative memory tests (immediate free recall, delayed free recall, cued recall, and recognition), several versions of each test (e.g., word lists 1, 2 and 3; and paragraphs 1, 2 and 3), and was conducted over two days. However, due to the complexity of the study's experimental design, only immediate and delayed recall of paragraph 3 on day 1 could be included as the memory outcome measure for the stress group. The memory outcome measure for the control group was immediate and delayed recall of paragraph 1, because it occurred before the stressor, and on the same day as retrieval of paragraph 3. None of the other groups had comparable control groups.

Smeets et al., (2009) also had two stress groups: *pre-stress learning* and *post-stress immediate learning*, as well as an appropriate control group- this group was exposed to the TSST, but only encoded the word list 2 hours after the TSST ended, thereby giving enough time for cortisol levels to have returned to baseline). This paper was therefore sub-divided into two separate studies: (i) Smeets et al., (2009) pre-stress learning group, who were exposed to the stressor at consolidation, and (ii) Smeets et al., (2009) post-stress immediate learning group, who were exposed to the stressor at encoding.

Stawski et al., (2009) had two word lists, and tested delayed recall and immediate recall of both word lists. However, word list 1 was encoded and immediately recalled prior to the stress manipulation, whereas delayed recall of list 1 was tested after the manipulation. This means that the stressor was in no way affecting memory

processing at immediate recall, and so only delayed recall for word list 1 was included. Word list 2, however, was encoded, immediately recalled, and then tested for delayed recall all after the stress manipulation. In this case, both immediate and delayed recall were taken into account for list 2.

Furthermore, separate *d*'s (effect sizes) were also calculated for studies that had more than one memory measure (e.g., immediate recall, delayed free recall, and recognition). For example, Espin et al., (2013) tested immediate free recall and delayed free recall, and so separate effect sizes were calculated for each of these memory measures.

References

Abercrombie, H. C., Kalin, N. H., Thurow, M. E., Rosenkraw, M. A., & Davidson, R. J. (2003). Cortisol variation in humans affects memory for emotionally laden and neutral information. *Behavioural neuroscience, 17,* 505-516.

Alderson, A. L., & Novack, T. A. (2002). Neurophysiological and clinical aspects of glucocorticoids and memory: A review. *Journal of Clinical and Experimental Neuropsychology, 24,* 335-355.

Alzheimer's Association. (2013). Alzheimer's disease facts and figures. *Alzheimer's Dementia, 9,* 208-245.

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

Beckner, V. E., Tucker, D. M., Delville, Y., & Mohr, D. C. (2006). Stress facilitates consolidation of verbal memory for a film but does not affect retrieval.

Boehringer, A., Schwabe, L., & Schachinger, H. (2010). A combination of high stress-induced tense and energetic arousal compensates for impairing effects of stress on memory retrieval in men. *Stress, 13,* 444-453.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science, 281,* 1185-1187.

Burke, H. M., & Davis, M. C., Otte, C., & Mohr, D. C. (2005). Depression and cortisol response to psychological stress: A meta-analysis. *Psychoneuroendocrinology, 30,* 846-856.

Cahill, L. (2003). Progress in neuro-psychopharmacology and biological psychiatry. *Stress Effects on Limbic Functions and Behaviour, 27,* 1235-1241.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Quervain, D. J., Aerni, A., Schelling, G., & Roozendaal, B. (2009). Glucocorticoids and the regulation of memory in health and disease. *Frontiers in Neuroendocrinology, 30,* 358–370.

De Quervain, D. J.-F., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience, 3,* 313-314.

De Quervain, D. J.-F., Henke, K., Aerni, A., Treyer, V., McGaugh, J. L., Berthold, T… Hock, C. (2003). Glucocorticoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe. *European Journal of Neuroscience, 17*, 1296-1302.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355-391.

Domes, G., Heinrichs, M., Reichwald, U., & Hautzinger, M. (2002). Hypothalamic-pituitary-adrenal axis reactivity to psychological stress and memory in middle-aged women: High responders exhibit enhanced declarative memory performance. *Psychoneuroendocrinology, 27,* 843-853.

Domes, G., Heinrichs, M., Rimmele, U., Reichwald, U., & Hautzinger, M. (2004). Acute stress impairs recognition for positive words- association with stress-induced cortisol secretion. *Stress, 7,* 173-181.

Eichenbaum, H. (2003). The hippocampus, episodic memory, declarative memory, spatial memory… where does it all come together? *International Congress Series, 1250*, 235-244.

Elzinga, B. M., Bakker, A., & Bremner, J. D. (2005). Stress-induced cortisol elevations are associated with impaired delay but not immediate recall. *Psychiatry Research, 134*, 211-223.

Elzinga, B. M., & Bremner, J. D. (2002). Are the neural substrates of memory the final common pathway in posttraumatic stress disorder (PTSD)? *Journal of Affective Disorders, 70*, 1-17.

Elzinga, B. M., & Roelofs, K. (2005). Cortisol-induced impairments of working memory require acute sympathetic activation. *Behavioural Neuroscience, 119,* 98-103.

Elzinga, B. M., Bakker, A., & Bremner. J. D. (2005). Stress-induced cortisol elevations are associated with impaired delayed, but not immediate recall. *Psychiatry Research, 134,* 211-223.

Espin, L., Almela, M., Hidalgo, V., Villada, C., Salvador, A., & Gomez-Amor, J. (2013). Acute pre-learning stress and declarative memory: Impact of sex, cortisol response and menstrual cycle phase. *Hormones and Behaviour, 63*, 759-765.

Field, A. P., & Gillet, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63* 665-694.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educaitonal Researcher, 5*, 3-8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Haberlandt, K. (1999). *Human memory. Exploration and application*. Boston, MA: Allyn and Bacon.

Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology, 30,* 771-784.

Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions.*

Jelicic, M., Geraerts, E., Merckelbach, H., & Guerrieri, R. (2004). Acute stress enhances memory for emotional words, but impairs memory for neutral words. *International Journal of Neuroscience, 114*, 1343-1351.

Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience, 10*, 459-466.

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: How does it work? *Trends in Cognitive Sciences, 10*, 152-158.

Kemeny, M. E. (2003). The psychobiology of stress. *Current Directions in Psychological Science, 12*, 124-129.

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine, 61*, 154-162.

Kirschbaum, C., Pirke, K-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test': A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76-81.

Kirschbaum, C., Pirke, K-M., & Hellhammer, D. H. (1995). Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. *Psychoneuroendocrinology, 20*, 509-514.

Kirschbaum, C., Prussner, J. C., Stone, A. A., Federenko, I., Gaab, J., Lintz, D… Hellhammer, D. H. (1995). Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men. *Psychosomatic Medicine, 57*, 468–474.

Kirschbaum, C., Wolf, O. T., May, M., Wippich, W., & Hellhammer, D. H. (1996). Stress- and treatment-induced elevation of cortisol levels associated with impaired declarative memory in healthy adults. *Life Science, 58*, 1475-1483.

Kirschbaum, C., Wüst, S., & Hellhammer, D. (1992). Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine, 54*, 648-657.

Kudielka, B. M., Hellhammer, D. H., & Wüst, S. (2009). Why do we respond so differently? Reviewing determinants of human salivary cortisol response to challenge. *Psychoneuroendocrinology, 34*, 2-18.

Kudielka, B. M., & Kirschbaum, C. (2005). Sex differences in HPA axis responses to stress: A review. *Biological Psychology, 69*, 113-132.

Kudielka, B. M., Schmidt-Reinwald, A. K., Hellhammer, D. H., & Kirschbaum, C. (1999). Psychological and endocrine responses to psychosocial stress and dexamethasone/corticotropin-releasing hormone in healthy postmenopausal women and young controls: The impact of age and a two-week estradiol treatment. *Neuroendocrinology, 70*, 422-430;

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience, 25*, 2977-2982.

Künzel, H. E., Binder, E. B., Nickel, T., Ising, M., Fuchs, B., Majer, M… Modell, S. (2003). Hypothalamic-pituitary-adrenocortical axis reactivity in acutely depressed psychiatric in-patients, measured by the Dex-CRH test. *Neuropsychopharmacology, 28,* 2169-2178.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Orlando, Fl: American Evaluation Association.

Luethi, M., Meier, B., & Sandi, C. (2009). Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioural Neuroscience, 2,* 1-9.

Lupien, S. J., Gillin, C. J., & Hauger, R, L. (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A dose-response study in humans. *Behavioral Neuroscience, 113,* 420-430.

Lupien, S.J., & Schramek, T. E. (2006). The differential effects of stress on meory consolidation and retrieval: A potential involvement of reconsolidation? Theoretical comment on Beckner et al. (2006). *Behavioural Neuroscience, 120,* 735-738.

Maheu, F. S., Collicutt, P., Rachel, Kornik., Moszkowski, R., & Lupien, S. J. (2005). The perfect time to be stressed: A different modulation of human memory by stress applied

in the morning or in the afternoon. *Progress in Neuropsychopharmacology and Biological Psychiatry, 29,* 1281-1288.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11,* 442-450.

Mason, J. W. (1968) A review of psychoendocrine research on the pituitary-adrenal cortical system. *Psychosomatic Medicine, 30,* 576--607.

McEwen, B. S., & Sapolsky, R. M. (1995). Stress and cognitive function. *Current Opinion in Neurobiology, 5,* 205-216.

Moayyedi, P. (2004). Meta-analysis: Can we mix apples and oranges? *The American Journal of Gastroenterology, 99,* 2297-2301.

Nater, U. M., Moor, C., Okere, U., Stallkamp, R., Martin, M., Ehlert, U., & Kliegel, M. (2007). Performance on a declartive memory task is better in high than low cortisol responders to psychosocial stress. *Psychoneuroendocrinology, 32,* 758-763.

Nicholson, S. A. (1989). Stimulatory effect of caffeine on the hypothalamo-pituitary-adrenocortical axis in the rat. *Journal of Endocrinology, 122,* 535–543.

Oei, N. Y. L., Everaerd, W. T. A. M., Elzinga, B. M., Van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress, 9,* 133-141.

Pickering, T. G., & Gerin, W. (1990). Cardiovascular reactivity in the laboratory and the role of behavioral factors in hypertension: A critical review. *Annals of Behvavioral Medicine, 12,* 3-16.

Rey, A. (1993). Psychological examination of traumatic encephalopathy (translation). *The Clinical Neuropsychologist, 7,* 3-21.

Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 803–814.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183-192.

Rosenthal, R., & Rosnow, R, L. (2008). *Essentials of behavioural research: Methods of behavioural research* 2rd ed. NY: McGrawHill.

Roozendaal, B. (2002). Stress and memory: Opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory, 78,* 578-595.

Sauro, M. D., Jorgensen, R. S., & Pedlow, C. T. (2003). Stress, glucocorticoids, and memory: A meta-analytic review. *Stress, 6,* 235-245.

Schöneshöfer, M., & Wagner, G. G. (1977). Sex differences in corticosteroids in man. *The Journal of Clinical Endocrinology & Metabolism, 45*, 814-817.

Schoofs, D., & Wolf, O. T. (2009). Stress and memory retrieval in women: No strong impairing effect during the luteal phase. *Behavioural Neuroscience, 123*, 547-554.

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O.T., & Oitzl, M.S. (2012). Stress effects on memory: An update and integration. *Neuroscience and Biobehavioural Reviews, 36, 1740-1749.*

Schwabe, L., & Wolf, O. T. (2010). Learning under stress impairs memory formation. *Neurobiology of Learning and Memory, 93,* 183-188.

Smeets, T., Giesbrecht, T., Jelicic, M., & Merckelbach, H. (2007). Context-dependent enhancement of declarative memory performance following acute psychosocial stress. *Biological Psychology, 76,* 116-123.

Smeets, T., Jelicic, M., & Merckelbach, H. (2006a). Stress-induced cortisol responses, sex differences, and false recollections in a DRM paradigm. *Biological Psychology, 72,* 164-172.

Smeets, T., Jelicic, M., & Merckelbach, H. (2006b). The effect of acute stress on memory depends on word valence. *International Journal of Psychophysiology, 62,* 30-37.

Smeets, T., Sijstermans, K., Gisjen, C., Peters, M., Jelicic, M., & Merckelbach, H. (2008). Acute consolidation stress enhances reality monitoring in healthy young adults. *Stress, 11*, 235-245.

Smyth, J. M., Ockenfels, M. C., Gorin, A. A., Catley, D., Porter, L. S., Kirschbaum, C… & Stone, A. A. (1997). Individual differences in the diurnal cycle of cortisol. *Psychoneuroendocrinology*, *22*, 89-105.

Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework* (NASA Technical Memorandum 212824). Moffett Field, CA: NASA Ames Research Center.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195-213.

Stawski, R. S., Sliwinski, M. J., & Smyth, J. M. (2009). The effects of an acute psychosocial stressor on episodic memory. *The European Journal of Cognitive Psychology, 21*, 897-918.

Taylor, E. M. (1959). *Psychological appraisal of children with cerebral defects*. Cambridge, MA: Harvard University Press, pp. 423–428.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381-391.

Tulving E. (1995). Organization of memory: Quo vadis? In *The Cognitive Neurosciences*, ed. MS Gazzaniga, pp. 839–47. Cambridge, MA: MIT Press.

Vingerhoets, A. J. J. M., & Perski, A. (1999). The psychobiology of stress. *Psychology in Medicine. Bohn Stafleu Van Loghum, Houten/Diegem*, 34-49.

Walker, M. P., Brakefield, T., Hobson, J. A., Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*, 616-620.

Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale-Revised: Manual*. San Antonio: Psychological Corporation.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121*, 331-354.

Wolf, O. T. (2003). HPA axis and memory. *Best Practice & Research Clinical Endocrinology and Metabolism, 17,* 287-299.

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research, 1293*, 142-154.

Wolf, O. T., Convit, A., McHugh, P. F., Kandil, E., Thorn, E. L., De Santi, S... de Leon, M.J. (2001). Cortisol differentially affects memory in young and elderly men. *Behavioral Neuroscience, 155,* 1002-1011.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory.