Children's True and False Memories of Valenced Material

By

Laura J. Pearce

Thesis submitted for the degree of Doctor of Philosophy

School of Psychology

University of Sussex

June 2017

Declaration

Chapter II: *Children's Internalizing Reactions to Scary TV: A Meta-Analysis*. The author contributions are as follows: Laura Pearce was responsible for all aspects of data collection including the literature search and contacting key authors, for coding the data and extracting effect sizes, and for writing the manuscript; Andy Field was responsible for the initial conception of the research, providing feedback on study design, providing the meta-analysis and sensitivity bias scripts, providing feedback on the manuscript for Chapter II and for collaborating on an alternative manuscript for publication.

Chapter III: *Children's True and False Memory for Media: Valence, Stimulus Modality and Congruency*; Chapter IV: *Effects of Emotional Context and Valence on Children's True and False Memories* and; Chapter V: *Depth of Processing and Valence Effects on Children's Emotional Memory* are written in the style of articles appropriate for *Experimental Child Psychology*.

The author contributions are as follows: Laura Pearce and Andy Field were collectively responsible for initial conception of the research. Laura Pearce was responsible for all aspects of data collection, data analysis and writing of the manuscripts; Andy Field was responsible for providing feedback on study design, analysis and manuscripts.

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree. However this thesis incorporates to the extent indicated below, material already submitted as part of required coursework for the degree of MRes in Psychological Methods, which was awarded by the University of Sussex.

A previous version of the meta-analysis presented in Chapter II was analysed and submitted as part of a module of work for the MRes qualification. A substantial amount of work was later carried out during the PhD by Laura Pearce, including expansion of the search terms, contacting additional researchers for any unpublished data, and using new methods of analysis. Chapter II presents the updated data and analysis in a different theoretical context, and has been rewritten in its entirety by Laura Pearce.

A paper based on the data and analysis in Chapter II has been published in *Human Communication Research* as:

Pearce, L. J., & Field, A. P. (2016). The impact of 'Scary' TV and film on children's internalizing emotions: A meta-analysis. *Human Communication Research, 42*(1), 98-121. doi:10.1111/hcre.12069.

Signature: Laura Pearce

Acknowledgements

Firstly, I would like to thank my Mum, my Dad, Jin, and Alison for their loving support throughout my postgraduate studies.

Secondly, a massive thanks to all of my friends from Sussex University who have believed in me, supported me, helped me through difficult times, and who I have celebrated achievements with: Becky Grist, Hazel Anderson, Kerry Fox, Georgia Leith, Fidelma Hanrahan, John Maule, Zoë Field, and Ellie Snape, amongst many others. Thanks to Alice Skelton and Jesse Bull for their assistance in coding data in Chapter III, Maxine Sherman and Andy Mealor for their advice with signal detection methodology and analysis, and Sam Berens for his technical advice in Chapter V.

To my other friends outside of Sussex University in no particular order: Beth Parry and James Parmenter for always sharing their home with me, Andy Hicks for lending his voice to my studies and always offering a new perspective, Sandy Gould for his unwavering patience, Rachel Schmitz for her wit, humour and all round sparkling personality, Phil Xavier for always bringing the rum, and Jesse Bull for inspiring, encouraging and motivating me through the final stages of this PhD. You are all wonderful, amazing people.

I would like to thank Robin Banerjee for his helpful feedback on Chapters III-V, the other members of the CATTLab for useful discussions, and the ESRC and School of Psychology for providing me with the opportunity to pursue this PhD. I am fortunate to have had supervision from Professor Andy Field, with his academic expertise, sense of humour and patience. Thank you for your guidance, feedback and encouragement over the past five years.

Thank you to all of the schools who participated: Brighton and Hove High School Junior School, Broadmere Primary Academy, Chesswood Junior School, City Whitehawk Academy, The Holy Family Catholic Primary School (Addlestone), The King Alfred School (London), Ripley CofE Primary School, Seaford Primary School, St John the Baptist Catholic Primary School (Brighton), St Paul’s CofE Nursery and Primary School (Brighton), West Byfleet Junior School, and West Rise Junior School. Last but not least, thank you to all the parents who provided consent, and to the hundreds of children who took part in my experiments.

Contents

Summary

There has been a rise in anxiety amongst typically developing children in recent years. Existing research has suggested a link between the increase of television viewing, and the increase in childhood anxiety. This thesis confirms the plausibility of this hypothesis; a meta-analysis found a small but consistent effect of viewing “scary” television on children’s internalizing responses.

Existing cognitive models of emotional processing in anxious individuals identify attention, interpretation, and memory preferences towards emotionally negative materials as potential mediators. Whilst attention and interpretation preferences have been well evidenced, the link between anxiety and negativity preferences in memory is more tenuous, particularly in typically developing children. A positive-negative asymmetry in memory is well established in adults; however, the extent to which children process and remember positive, negative and neutral stimuli differentially is relatively unexplored. The Deese–Roediger–McDermott (DRM) paradigm has been utilised to allow analysis of memory accuracy and response bias.

There were several key findings within this thesis. In line with previous research, a preference away from negatively valenced material was found when simple word list stimuli was used. When narratives were used as a richer source of material, memory performance was greater for negatively, than positively valenced stimuli. However, when two sources differing on richness of information (visual vs narrative stimuli), and valence (positive vs negative) were presented simultaneously, the modality effect became dominant; the richer source of material was discriminated with higher accuracy, regardless of valence. When this effect was followed up, no evidence of a mood interaction within emotional memory was found. Semantic elaboration was explored as a potential mechanism behind valence effects in memory. However, no positive findings were identified. Age, gender and trait anxiety did not reliably moderate valence effects on memory.

This thesis adds to the small body of knowledge focusing on children’s emotional memory, particularly by including response bias analyses. It highlights the complex nature of emotional processing in children and some of the factors contributing to accuracy. Further research should explore in greater depth how valence effects differ for various types of stimuli, and under which circumstances these effects can be overridden. Mechanisms behind these valence effects are also yet to be unpicked.

Chapter I: Overview

Introductory Remarks

Global levels of childhood anxiety are on the rise. A variety of clinical risk factors can increase the likelihood of developing internalizing disorders such as depression and anxiety. Early adverse experiences including child maltreatment (Famularo, Kinscherff, & Fenton, 1992; Margolin & Gordis, 2000), or being victimised or bullied (Bond, Carlin, Thomas, Rubin, & Patton, 2001; Troy & Sroufe, 1987) are some such examples. Rates of child anxiety also increased by around 1 standard deviation in typically developing children between 1952 and 1993, with the average American child in the 1980s reporting more anxiety than child psychiatric patients in the 1950s (Twenge, 2000). Twenge proposes that environmental influences such as decreased social connectedness and a perception of increased overall environmental threat have led to this rise. However, many models of anxiety postulate that disorders are caused by, or linked to differences in cognitive processing of emotional information (for a review see Hadwin & Field, 2010). Whilst literature on emotional preferences within attention and interpretation for clinically anxious and non-anxious individuals is relatively well established, research surrounding memory effects is less clear cut. This thesis aims to: (1) evaluate whether environmental influences (specifically “scary” TV) do affect children’s internalizing symptoms, such as anxiety; (2) to explore whether trait anxiety is related to memory performance for valenced material, within typically developing children; and (3) to explore memory for different types of valenced stimuli, within typically developing children. Potential moderators of emotional memory preferences are also explored.

Why Study Emotion in Children?

Internalizing (emotional) disorders are highly prevalent across children in Great Britain. Internalizing disorders are one of two broad psychopathology classes; the other being externalizing (behavioural) disorders. Internalizing disorders signify a core disturbance in intropunitive emotions and moods (e.g. sorrow, guilt, fear, and worry), whereas externalizing problems are characterized by behaviours that are harmful and disruptive to others (Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). The *Diagnostic and Statistical Manual of Mental Disorders (DSM-V*) categorizes internalizing disorders as those with prominent anxiety, depressive, and somatic symptoms, characterized by depressed mood, anxiety, and related physiological and cognitive symptoms (American Psychiatric Association [APA], 2013). Anxiety and depressive disorders are the primary internalizing disorders, although other disorders such as trauma related disorders, sleep-wake disorders, obsessive-compulsive disorders, and attention-deficit/ hyperactivity related disorders can also have internalizing components.

A 2004 Office of National Statistics (ONS) survey found a prevalence rate of 3.7% for emotional disorders in 5-16 year olds (Green, McGinnity, Meltzer, Ford, & Goodman, 2005). Prevalence rates were 0.9% for depression, and 3.3% for anxiety disorders, of which specific phobia (0.8%) and generalized anxiety (0.8%) were most commonly reported. Research papers generally report much higher rates of emotional disorders, indicating that many children and adolescents may not have received a formal diagnosis, despite meeting criteria as set out by the *DSM-V* (APA, 2013) or *International Statistical Classification of Diseases and Related Health Problems* (*ICD-10*; World Health Organization [WHO], 2010). A review by Beesdo, Knappe, and Pine (2009) reported prevalence rates for anxiety disorders of between 3.1% (Gau, Chong, Chen, & Cheng, 2005) and 29.9% (Woodward & Fergusson, 2001), concluding that the lifetime prevalence of "any anxiety disorder" in studies with children or adolescents is about 15% to 20% (Beesdo et al., 2009). Higher levels of depression are also reported, at around 2% in children and 5-8% in adolescents (Birmaher et al., 1996; Jellinek & Snyder, 1998).

Children and adolescents with anxiety disorders are at higher risk of developing comorbid psychopathologies (Kendall, Brady, & Verduin, 2001), including depression (Brady & Kendall, 1992; Kovacs, Gatsonis, Paulauskas, & Richards, 1989), externalizing disorders (for a review see Russo and Beidel, 1994), and substance use disorders (Lewinsohn, Rohde, & Seeley, 1995). Anxiety is also believed to be a contributing factor in gastrointestinal disease, cardiovascular disease and chronic pain, although this is more common in adults (Roy-Byrne et al., 2008).

Levels of anxiety are also increasing in non-clinical samples of children. The average American child in the 1980s reported more anxiety than child psychiatric patients in the 1950s (Twenge, 2000). Subclinical manifestations of anxiety disorders have been identified in 49.0% of children interviewed (Muris, Merckelbach, Mayer, & Prins, 2000). Children with subclinical levels of anxiety are at increased risk of developing anxiety disorders at a clinical level (Donovan & Spence, 2000; Rapee, Kennedy, Ingram, Edwards, & Sweeney, 2005). The amount of anxiety that typically developing children and adolescents currently report is at such a high level that it is vital to understand the underlying causes and maintenance mechanisms to ultimately provide effective, targeted support.

Anxiety in children has been linked to poorer school performance (Ashcraft, 2002; Crozier & Hostettler, 2003; Mazzone et al., 2007; Wood, 2006), poorer social functioning (Ginsburg, Greca, & Silverman, 1998; Greca & Lopez, 1998; Wood, 2006), and poorer family functioning (Hughes, Hedtke, & Kendall, 2008; Siqueland, Kendall,

& Steinberg, 1996). The majority of these studies note that it is unclear whether these factors are causes, consequences or maintenance factors in the development of childhood anxiety.

Why Study Children's Emotional Responses to Scary TV?

A multitude of factors have been identified which may be contributing to the rising levels of childhood anxiety. For example, Gray (2011) cites the decline of play as a major factor in the rise of childhood psychopathologies. Childline reported in 2016 that the number of children and young people needing counselling in relation to online bullying had increased by 88% over the previous five years (NSPCC, 2016). A report commissioned by the National Union of Teacher (NUT) also concludes that rising accountability measures (including public exams such as SATs) effectively turn schools into 'exam factories', which impacts on the pressure children and young people feel, increasing their levels of anxiety, disaffection and mental health problems (Hutchings, 2015).

Additionally, Twenge (2000) identified two social indicators which have changed since the 1950s and are significantly linked to an increase in anxiety: social connectedness, and a perception of overall environmental threat. The rise in television (TV) consumption has been cited as a possible factor behind a lack of social connectedness and increase in perceived overall threat. In just a 10 year span from 1999–2009, the average amount of TV sets per household has increased from 3.1 to 3.8. The average amount of TV children watched also increased from 3 hr 47 min, to 4 hr 29 min per day (Rideout, Foehr, & Roberts, 2010). A rise in the use of technology, such as TV was also cited by Gray (2011) as a factor in the decline of play; further, TV is a source of information regarding other factors which might increase anxiety (such as news related to school exam results). Due to the cross cutting nature of television in children's lives, the effect of viewing of negative (scary) TV on internalizing emotions, including anxiety has been selected as one factor to explore in more depth.

Why Study Children's Emotional Memory?

With levels of both clinical and sub-clinical anxiety on the rise in childhood and adolescence, it is vital to explore not only which environmental factors (such as exposure to TV) can affect anxiety, but potential mechanisms by which exposure to negative materials may be linked with anxiety. Traditional cognitive models of anxiety are built on a schema-based information processing perspective (Beck, 1985). In Beck's early model, anxiety was thought to result when a situation arose, was encoded, and subsequently interpreted as fitting within a threat schema. The term '*schema*' appears to be used inconsistently throughout literature, but is often described as a knowledge framework within memory that selects and actively modifies experience in order to arrive at a coherent, unified, expectation-confirming and knowledge-consistent representation of an experience (Alba & Hasher, 1983). Schema theory commonly includes four stages: selection, abstraction, interpretation, and integration of stimuli into existing memory schema.

More recently, theories have described various distortions apparent in various steps in cognitive processing in children exhibiting anxiety (see Muris & Field, 2008 for a review). A cognitive distortion refers to a systematic pattern of deviation from the norm, or rationality in judgment whereby inferences about other people and situations may be drawn in an illogical fashion. Muris and Field (2008) propose an integrative theoretical model of cognitive distortions on the processing of threat-related information, such as scary TV; processing patterns found in anxious individuals, but not non-anxious individuals. These are hypothesized to play a role in the maintenance and/or exacerbation of childhood anxiety (see Figure 1).

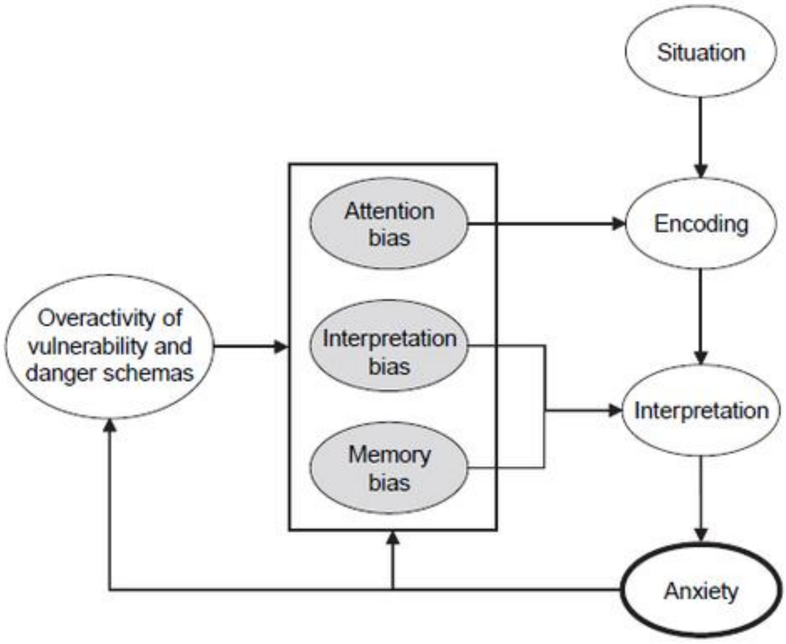


Figure 1. Theoretical model showing the influence of cognitive distortions on the processing of threat-related information, which is hypothesized to play a role in the maintenance and/or exacerbation of childhood anxiety. Figure from Muris and Field, 2008.

Within this model, anxious children are thought to have overactive schemas involving the themes of vulnerability and danger. When confronted with potential threat, novelty, or ambiguity, schemas strongly guide the processing of information and chronically focus resources on threat-relevant information in three ways: (1) increased attention towards threatening stimuli compared to non-threatening stimuli, not seen in non-anxious individuals; (2) increased propensity for negative interpretations of ambiguous situations, compared to non-anxious individuals; and (3) a tendency to selectively recall memories congruent with the emotional state of anxiety, not evident in non-anxious individuals.

Within memory literature, this type of cognitive distortion has been referred to inconsistently as an '*emotional enhancement effect*', '*negativity effect*', and '*negativity* *bias'* (see Murphy & Isaacowitz, 2008 for a review). These terms have been used to refer to a variety of outcomes: (1) where negative material is remembered at a significantly higher rate, or more accurately than neutral material; (2) where negative material is remembered at a significantly higher rate, or more accurately than positive material; (3) where one group of participants remember negative material at a significantly higher rate than another group; and (4) where negative material is remembered at the same rate between groups or conditions, but the rate of remembering positive material decreases. Where this effect is towards positive material, it is known as a positivity bias, effect, or enhancement effect. In line with the approach taken by Murphy and Isaacowitz, the term '*negativity preference*' will be used throughout the thesis, when negative material is recalled or recognized (or attended to) at a significantly higher rate than neutral or baseline material, and '*positivity preference*' will be used when positive material is recalled or recognized (or attended to) at a significantly higher rate than neutral or baseline material. Where differences between positive and negative performance are highlighted, these will not be referred to as a 'preference' because there is no baseline comparison.

Emotional Preferences in Clinically Anxious Individuals

Within the context of anxiety, an attentional preference for threat related material over neutral material was reliably demonstrated in anxious adults (k = 101. d = 0.45), and children (k = 11, d = 0.50) using a meta-analysis (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007). This effect held across different experimental paradigms and under a variety of experimental conditions. Non-anxious individuals did not exhibit this effect (k = 87, d = –0.01). A more recent meta-analysis found a similar effect of preference for threatening, compared to neutral material in anxious children (k = 38, d = 0.54) (Dudeney, Sharpe, & Hunt, 2015). The effect in non-anxious children (d = 0.15) was much smaller compared with non-anxious adults, but was still evident.

When examining memory preferences in adults, a review categorized by anxiety disorder found some support for a preference towards threatening material in implicit memory across all anxiety types, but preferences within explicit memory were mixed depending on disorder type (Coles & Heimberg, 2002). A preference towards threat related information in explicit memory was seen in individuals with panic disorder across most studies that used deep encoding procedures. The number of studies looking at posttraumatic stress disorder (PTSD) and obsessive compulsive disorder (OCD) are small, but offer some support, whilst studies including people with social phobia and generalized anxiety disorder (GAD) found little support for preferences towards threatening materials. Further, a meta-analysis of 165 studies found no overall relationship between anxiety and implicit memory, and that high-anxious adults do not differ from low-anxious adults in the selective recognition of threat-related information for hit rates, false alarms, or corrected hit rates (discrimination) (Mitte, 2008). However, high-anxious adults did exhibit better recall for threatening material and poorer recall for positive material compared with low-anxious adults. The authors note that effect sizes are small, with the magnitude of the effect depending on experimental study procedures like the encoding procedure or retention interval.

Research into memory preferences in anxious children is more limited. It appears that there is a trend for an increase in anxiety predicting lower memory performance for positive compared to neutral stimuli, across several task types. Children with PTSD were found to recall significantly lower proportions of positive and neutral words than a control group, but the same proportion of negative words (Moradi, Taghavi, Neshat-Doost, Yule, & Dalgleish, 2000). Similarly, children with high social anxiety recalled less positive self-referent words than low socially anxious counterparts (Vassilopoulos, 2012). Daleiden (1998) had more mixed findings; an increase in anxiety levels predicted a positivity preference on procedural memory tasks, a negativity preference during conceptual but not perceptual memory tasks, and a preference away from positive compared to neutral information on declarative memory tasks.

In addition to the lack of research surrounding emotional memory preferences and anxiety in children, there is a limited amount of research into emotional memory preferences within the typically developing child population. Because anxiety is becoming increasingly prevalent in the typically developing population, it is important to understand whether the preferences that exist in clinically anxious adults, and to some extent in anxious children are atypical, or merely part of typical development. The remainder of this introduction initially introduces the theories and associated methodologies used to investigate recognition and recall memory, to explain and defend the methodologies used within the experiments of this thesis. The introduction then moves on to discuss what we already know about emotional memory in typically developing adults and children. Finally, the introduction will provide an overview of the remaining chapters of the thesis.

An Introduction to Valenced Stimuli

When investigating emotional memory, the appropriate stimuli must be selected. Emotional stimuli can vary on two dimensions; arousal and valence. In a laboratory setting, valenced stimuli typically include words or images that have been judged as negatively or positively valenced, using affective norms gained from large populations (Norris, Larsen, Crawford, & Cacioppo, 2011). There are many word, picture and sound banks from which researchers draw their stimuli, including: the Affective Norms for English Words (ANEW) (Bradley & Lang, 1999); AFINN (Nielson, 2011); The Berlin

Affective Word List Reloaded (BAWL-R) (Võ et al., 2009); The Cornell/Cortland Emotion lists (Brainerd, Toglia, Reyna, & Stahl, 2008); the International Affective Picture System (IAPS) (Center for the Study of Emotion and Attention (CSEA-NIMH), 1995); the Geneva Affective Picture Database (GAPED) (Dan-Glauser & Scherer, 2011); and the International Affective Digitized Sounds (IADS) bank (Bradley & Lang, 1999).

However, emotional stimuli can also vary on strength of arousal (how exciting or calming a stimulus is) (Adelman & Estes, 2013). Typically, more strongly valenced stimuli are also more strongly arousing (Bradley & Lang, 1999), and stimuli with a negative valence are rated higher in arousal than stimuli with an equally positive valence (for example, Citron, Weekes, & Ferstl, 2014). There is ongoing debate surrounding the relative contributions of each to the effect that emotional material has on memory. Historically, much research has pointed towards arousal being the sole contributor (Hamann, 2001; Mather, 2007; Phelps, 2005), meaning that both positive and negative stimuli would enhance memory if suitably high on arousal. Other research suggests valence affects memory performance independent of arousal (Kensinger & Corkin, 2003, 2004), or has an effect on the processing of different aspects of the stimuli within memory (Kensinger, 2009; Kensinger & Schacter, 2006, 2008). These studies revealed that the lateral prefrontal cortex responded differentially to negative items, whereas the medial prefrontal cortex was more engaged during the processing of positive pictures. This corresponded to increased perceptual processing for negative items, and increased conceptual processing for positive items, again indicating that negative and positive material could enhance memory under specific circumstances. Additionally, some existing evidence suggests a curve-linear relationship between valence and arousal measures (Söderholm, Häyry, Laine, & Karrasch, 2013), and an interaction effect has been reported between valence and arousal on word recognition tasks (Citron, Weekes, & Ferstl, 2014). It is therefore important to ensure that selected emotional stimuli are matched not only on intensity of valence, but levels of arousal during laboratory studies.

The Cornell/Cortland word lists (Brainerd, Toglia, et al., 2008) have been used in Chapters IV and V of this thesis, because they are split not only by positive, negative and neutral valence, but also by low and high arousal within each valence category. Word lists of low and high arousal were selected for each valence condition. Ratings from large populations were also used to produce norms for arousal and valence strength for each word. Valence and arousal norms were taken from the Warriner, Kuperman, and Brysbaert norms for 13,915 English lemmas (2013) to check the main manipulation of valence, and to investigate arousal as a potential confounding variable. Word lists can also vary on many other characteristics that should also be held constant where possible, or considered when interpreting results of a memory study. Word frequency, familiarity, word length, neighbourhood size, imageability and age of acquisition have all been shown to affect response time and accuracy in recognition tasks (Atkinson & Juola, 1971; Brown & Watson, 1987; Carroll & White, 1973; James, 1975; Luce & Pisoni, 1998; Morrison & Ellis, 1995; Whaley, 1978). Scores across these features were obtained using N-Watch (Davis, 2005) for all words presented in Chapters IV and V. Analysis of the word list properties is presented in the results section of Chapter IV, with further details presented in supplementary materials.

Memory Methodology

Once a suitable stimulus set has been chosen, either recall or recognition tasks may be employed to study emotional memory. In their most basic form, recall refers to the retrieval of encoded and stored information in the absence of stimuli, whereas recognition refers to the ability to judge whether presented material matches previously presented material (Medina, 2008). Recognition memory is therefore a subtype of declarative memory. There are several subtypes of recognition and recall tasks, with different strengths and limitations, allowing different components of memory to be studied.

Recognition Memory Theory

Signal detection theory. Signal detection theory (SDT) is applied when two stimulus types must be discriminated (Stanislaw & Todorov, 1999). Participants are required to discriminate between '*signals'* (previously presented stimuli) and '*noise*' (stimuli not previously presented). SDT is based on the assumption that there is no differentiation between stages of simply knowing a signal has been presented before, and actually being able to recollect the experience. Recognition occurs when a value of a decision variable is sufficiently high to meet the criterion of a signal. SDT has been used as a basis to analyse the recognition task results throughout the experimental chapters of this thesis. A brief overview of the types of tasks and calculations used, and a critique of the approach are provided here. For a fuller review of the overarching theory see Green and Moses (1966) and Green and Swets (1966).

Recognition Memory Methodology

Yes/No (Old/New) task. Within memory studies a Yes/No paradigm is most commonly utilized. Experiments in Chapters IV and V employed a Yes/No task. In the Yes/No paradigm, stimuli (e.g. lists of words presented verbally or visually) are presented in a learning phase. In the subsequent test phase participants are shown some, or all of the original stimuli (targets) and some distractor stimuli (noise). Participants can either be asked to indicate whether they have seen the stimulus before (Yes/No), or whether the stimulus is Old/New (Stanislaw & Todorov, 1999). As long as participants are using a relevant decision criterion, and are able to distinguish signals from noise, then signals should seem more familiar than noise in memory tasks.

Responses are coded into one of four categories: a *hit* (H) corresponds to a ‘yes’ for a target, whereas a *miss* is a ‘no’ response to a target, a *false alarm* (F) is a ‘yes’ response to a distractor, and a *correct rejection* is a ‘no’ response to a distractor. Within the SDT model, a hit signifies a *remember* response. These responses are used to calculate measures such as sensitivity or discriminability (*d*’) and response bias, or criterion *C* (*C*). Sensitivity shows whether respondents are able to discriminate signals from noise significantly better than at chance level. The larger the value of *d*’, the better discrimination is between signals and noise. The following formula is commonly used (Macmillan, 1993), effectively subtracting the *z* score of false alarm rates (the probability of responding yes to distractors) from the *z* score of hit rates (the probability of responding yes to targets) – see Eq. 1.

$$d' = \Phi^{-1}(H) - \Phi^{-1}(F) \qquad \text{Eq. 1}$$

When H or F rates equal 0 or 1, the use of the Φ^{-1} function is problematic because the corresponding *z* scores are $-\infty$ and $+\infty$, respectively. To combat this problem, data can be combined from several participants before calculating H and F rates (Macmillan & Kaplan, 1985). This process relies on participants exhibiting comparable sensitivity and response bias, and complicates statistical analysis, and so it is not a preferred method. A second approach is to adjust only the extreme measures themselves by replacing scores of 0 with $0.5/n$ and scores of 1 with $(n - 0.5)/n$, where n is the number of trials (Macmillan & Kaplan, 1985). However, this adjustment results in a statistically biased sensitivity measure (Miller, 1996). A third option is the *loglinear* approach; to adjust all the scores by the same amount to retain the relationship between them, and therefore preserve the unbiased sensitivity measure (Hautus, 1995). This approach entails the addition of 0.5 to all H and F scores, and 1.0 to the number of targets and number of distractor trials responded to, prior to calculating H and F rates. This approach has been used throughout the thesis.

Discrimination is unaffected by response bias and is, therefore, a pure measure of sensitivity when the two *d*' assumptions are met: (1) both signal and noise distributions must be normal; and (2) the standard deviation for both signal and noise distributions must be equal. These assumptions cannot be tested by the Yes/No task, and are arguably not always tenable (Swets, 1986). Where this is suspected, nonparametric measures such as *A'* may be used (Pollack & Norman, 1964). For further discussion of nonparametric measures and formulae see Stanislaw & Todorov (1999). A *d*' measure was calculated throughout this thesis because the calculations are statistically more straightforward, and there were no concerns that signal and noise distributions, and their *SDs* would be incomparable. These assumptions could not be directly tested in Chapters IV and V however, because a Yes/No task was employed.

Criterion *C* scores show the general tendency to respond *yes* (or *no*) regardless of whether the stimulus is a signal or distractor. The standard formula for *C* is shown in Eq. 2 (Macmillan, 1993).

$$C = \frac{\Phi^{-1}(H) + \Phi^{-1}(F)}{2} \qquad \text{Eq. 2}$$

Figure 2 illustrates the relationship between *d*' and *C*. It shows an example where a decision criterion is arbitrarily set at 0.5. The signal and noise distributions show the values realized by the decision variable across signal and noise trials, respectively.

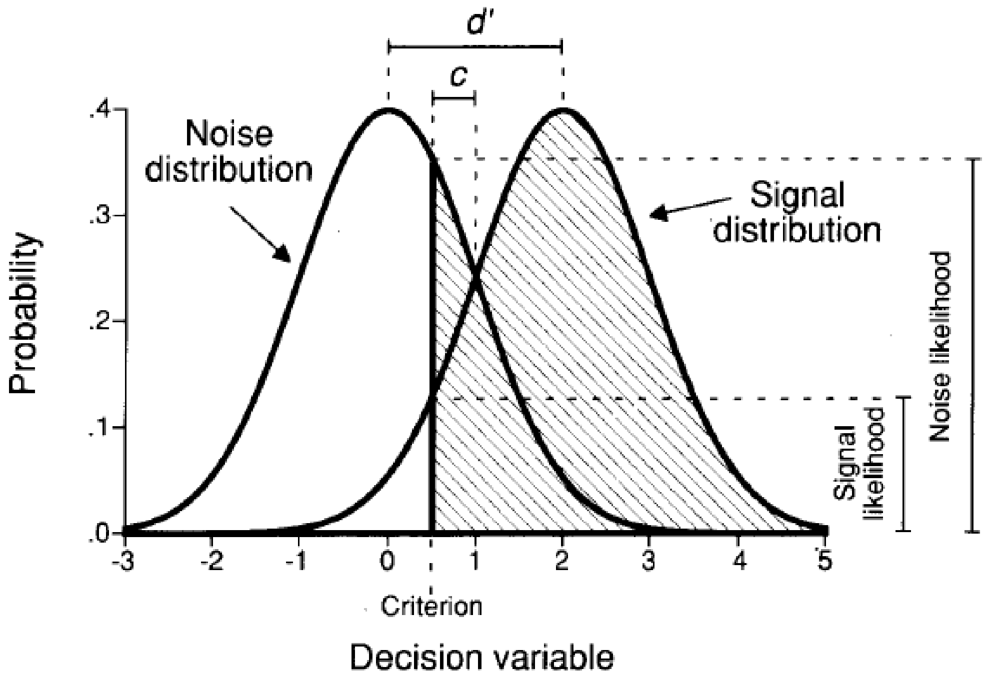


Figure 2. Distribution of the decision variable across noise and signal trials, showing *d' and C*. Figure from Stanislaw and Todorov (1999).

When a stimulus is presented and its value exceeds the 0.5 criterion threshold, it is classed as a signal. This means all shaded areas in the graph (about 30% of noise stimuli and 90% of signal stimuli) are classified with a 'yes' response. The hit rate equals the proportion of the signal distribution that exceeds the criterion (approximately 0.90), whereas the false alarm rate equals the proportion of the noise distribution that exceeds the criterion (approximately 0.30). The neutral point is located where the signal and noise distributions cross over (at 1.0). *C* is defined as the distance between the criterion and neutral point, where neither response is favoured. When *C* equals 0, the criterion will be located at the neutral point and there is no response bias. When a 'yes' response is more likely than a 'no' response (such as in Figure 2), the false alarm rate will exceed the miss rate and *C* becomes a negative value. The criterion appears to the left of the neutral point. When it varies significantly in this direction, it is classified as liberal response bias. If the decision criterion is set to a lower figure, it will become even more liberal, because more signals and distractors will be accepted. If the criterion is set higher than the neutral point then *C* will become positive; more stimuli would be rejected and a conservative response bias would be shown. This occurs when the miss rate exceeds the false alarm rate.

There are several methods for running SDT analysis. First, *d*' can be calculated using tabular methods (Elliott, 1964; Freeman, 1973), although this is noted to result in poor accuracy (Stanislaw & Todorov, 1999). Specialized SDT software has been created for Yes/No tasks such as the BASIC programme (Brophy, 1986) and the Pascal programme (Macmillan & Creelman, 2004). Source code such as RSCORE4 is also available for analysis of ratings task data (Dorfman, 1982). The simplest method to use without compromising accuracy is general purpose software with access to the Φ^{-1} function. MS Excel was therefore used throughout this thesis to compute measures of *d*' and *C* using the NORMSINV (Φ^{-}) function.

Rating tasks. The overall stimulus presentation and test phases for rating tasks are identical to the Yes/No procedure, with the exception that a ratings scale is used, rather than a dichotomous *yes* or *no* response. A signal detection rating task was used in the two experiments in Chapter III, rather than a Yes/No task. A rating task was chosen because the stimuli were more complex than simple word lists and more likely to lead to varying levels of certainty. A memory ratings task therefore allowed for a graded level of recognition response (Stanislaw & Todorov, 1999).

There is considerable debate around whether Likert scales should include a mid-point, based on both methodological and epistemological arguments. Some researchers argue that including a mid-point results in a high tendency to choose mid-point options, leading to a falsely elevated reliability score on the measure (Weems & Onwuegbuzie, 2001). However, primary school age children are no more likely to tend towards the centre of the scale when administered a 5-, compared to 4-point scale (Adelson & McCoach, 2010). In their study the mean, variance, and residuals were equal for both scales, but the 5-point scale also resulted in a higher reliability statistic compared to the 4-point scale, suggesting it is a better choice methodologically. From an epistemological perspective, various authors have noted that the true meaning of a mid-point is often unknown because participants will interpret the midpoint differently. For example, Raaijmakers, Van Hoof, Hart, Verbogt, and Vollebergh (2000) noted that the midpoint could relate to either neutral attitudes, or an *undecided/ don't know* response. If a mid-point is included, it must be made very clear to participants exactly what it represents. Likert-type scales can also be conceptually difficult for children to understand, and should be used with caution; formats with words are easier to understand than numbers (Mellor & Moore, 2013). A 5-point scale was therefore chosen for Experiments 1 and 2 to include a midpoint. Each point on the scale was labelled, and it was made clear to children that the mid-point represented '*I don't know*'.

Typically, ratings are used to determine receiver operating characteristic (ROC) curve points on a graph during analysis. However, to allow comparisons between results in Chapter III, IV, and V, scores of 1 (*sure this is false*) and 2 (*think this is false*) were grouped together as a *no* response, and scores of 4 (*think this is true*) and 5 (*sure this is true*) as a *yes* response, in Chapter III. Scores of 3 (*don't know*) have been treated as non-responses/missing data. The H and F rates have been calculated using the total number of items responded to, excluding non-responses.

Further analysis could also be conducted with this data to calculate meta-memory scores; the efficacy with which observers' confidence ratings discriminate between their own correct and incorrect stimulus classifications (Maniscalco & Lau, 2012). Whilst this is a little-explored area and would be of value to investigate further, it was not central to the aims of this thesis and so has not been reported.

Recall Memory Theory

Generate-recognize theory. Watkins and Gardiner (1979) provide an overview of a variety of two-stage recall models which all share core features. These models are based on memory-trace theory, whereby distinct concepts or ideas are stored as a finite number of nodes or atoms in a network (Anderson & Bower, 1972). This is one of the key theories of memory recall. A search is first conducted to identify associated pathways and generate potential items for recall. Recognition is then carried out to identify whether the item has an appropriate ‘tag’ assigned to it. Generate-recognize theories are therefore more in line with SDT, whereby recognition only occurs when value of a decision variable is sufficiently high to meet the criterion of a signal (Stanislaw & Todorov, 1999). This means that recall can fail at two stages, either the items are not generated initially, or they can be assessed incorrectly once generated. Traces can be strengthened by various aspects of stimuli such as valence, arousal, relatedness, and distinctiveness (Talmi, Luk, McGarry, & Moscovitch, 2007a; Talmi & Moscovitch, 2004; Talmi, Schimmack, Paterson, & Moscovitch, 2007). Emotional items tend to receive higher valence and arousal scores than neutral items. Emotional items are assumed to naturally be more related because they share categorical and thematic relatedness (Talmi et al., 2007a). Emotional items are also assumed to be more distinct in ‘absolute’ terms because they share a limited overlap of features in long term memory (LTM), and are more distinct in ‘relative’ terms because there is less overlap of items in the active conceptual framework in working memory (Schmidt, 1991). These features allow traces for emotional items to be strengthened during encoding.

Recall Memory Methodology

Free recall. In free recall tasks, participants are presented with stimuli and then asked to repeat back what was presented. As with recognition tasks, free recall can be initiated immediately after all stimuli have been presented (immediate recall) or after a delay, typically including a distractor task (delayed recall). Recall stimuli often consist of lists of words (e.g. Howe, 2007; Howe, Candel, Otgaar, Malone, & Wimmer, 2010; Howe & Derbish, 2010). Although word lists are typically used, participants may be asked to recall passages of text (Bartlett, 1932), images (Marks, 1973), or life events (Casey, Masuda, & Holmes, 1967). A delayed free recall task was employed in addition to a recognition task in Experiments 1 and 2.

Primacy and Recency Effects

Commonly, primacy and recency effects are observed within recall and recognition tasks (Oberauer, 2002). Items presented towards the beginning and end, respectively, of stimulus presentation have a higher probability of being remembered than those in the middle of the list. Figure 3 shows this effect for recall tasks with a 10 and 30 s delay between stimulus presentation and initiation of the recall task.

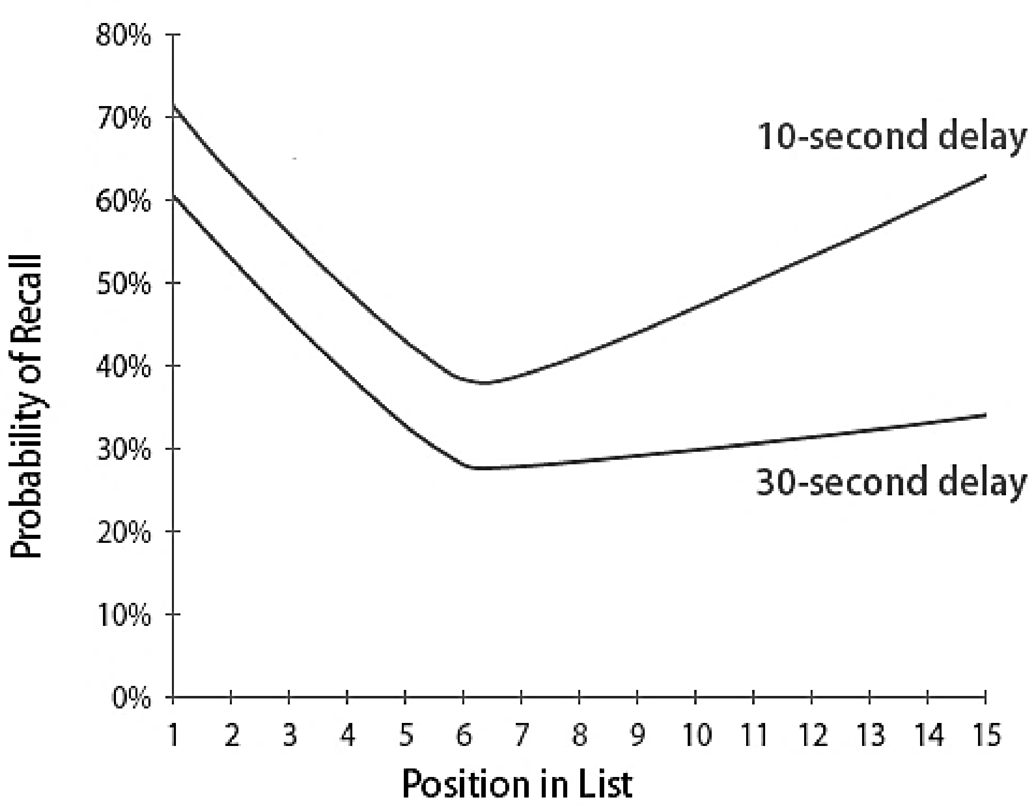


Figure 3. The primacy and recency effect in probability of recall, by length of delay between stimulus presentation and initiation of recall task. Figure from Universal Principles of Design (Lidwell, Holden, & Butler, 2003).

The primacy effect is thought to reflect an advantage in processing for earlier items, whereby they are rehearsed for a longer time, resulting in a transfer from short term to long term memory (Atkinson & Shiffrin, 1968, Fischler, Rundus, & Atkinson, 1970). The recency effect is thought to occur because later items are still being actively held in working memory and can be easily retrieved. Evidence for this comes from studies manipulating the length of the delay before recall. As the delay increases, recall of all items decreases, but particularly for those presented near the end of the list. This is illustrated in Figure 3. This effect is assumed to arise because the items have not been activated whilst still in working memory, but have also not been rehearsed due to the distraction task; which would result in them being discarded (Tan & Ward, 2000). Delayed free recall tasks are employed in Experiments 1 and 2, with the use of a distractor task in Experiments 1, 2, and 4 to ameliorate the recency, and to a lesser extent, the primacy effects.

False Memory: The Deese–Roediger–McDermott (DRM) Paradigm.

The DRM paradigm was pioneered by Deese (1959) and extended upon by Roediger and McDermott (1995) to explore false memory. Initially, Deese presented participants with 36 lists, each comprising 12 words from Kent-Rosanoff word-association lists (Russell & Jenkins, 1954). The 12 words on each list were all semantically related to a key word (*critical lure*), which was not presented to participants. Participants completed an immediate free recall task after each list. The number of times the critical lures were recalled was noted as a measure of false memory: a memory of an event that never occurred. False memory rates varied between 0% intrusion rate for the *butterfly*, and 44% intrusion rate for the *sleep* critical lure.

Roediger and McDermott (1995) expanded upon this by using the six lists Deese had identified as most likely to lead to false memory recall, and employing a recognition task at the end of the procedure. The recognition task contained 12 targets (2 from each list) and 30 distractors. Of the distractors, 6 were the critical lures from each list, 12 were *non-critical lures*, or items weakly related to the targets (2 items from below position 13 on each of the original lists, which had not been previously presented) and 12 were non-associated items. The recognition task employed a 4-point ratings scale from 1 (*sure the item is new*) to 4 (*sure the item is old*), with no midpoint for *don't know/can't remember*.

DRM studies have been used to explore how emotional processing changes with development, including across childhood. Typically, accuracy increases with age as audiospatial and visuospatial working memory capacity increase (Thomason et al., 2009; Vuontela et al., 2003). However, one exception to standard age improvements is when false information is presented that reflects a prototypical exemplar of a well-studied topic (critical lure). Older children and adults tend to erroneously remember these items as having been presented previously, with this effect increasing with age. These “developmental reversals” have been consistently found in many DRM studies, and are believed to be a result of age-related increases in associative connections among related information and increased reliance on gist (as opposed to surface) level processing of information while encoding and retrieving information. Sometimes, they can also lead to decreases in net accuracy, although this is not always found. For a review of the data and theories of developmental reversals in false memory see Brainerd, Reyna and Ceci (2008). The DRM procedure has been utilised in Chapters IV and V using fewer, and shorter lists than Deese (1959) and Roediger and McDermott (1995), respectively, to account for children’s lower working memory capacity.

Interpreting Memory Effects

Power. Power analysis is typically recommended in experimental studies (Marszalek, Barber, Kohlhart, & Holmes, 2011). Power is the probability of detecting an effect, if the effect actually exists. It is therefore the probability of finding a difference between two or more conditions, where there is a ‘*true*’ effect. Higher powered studies have higher precision to detect genuine effects. Power, effect size, sample size and alpha are related such that each is a function of the other three. In other words, if three of these values are fixed, the fourth is completely determined (Cohen, 1988). Typically, power analysis is conducted before collecting data, to determine what sample size is needed to detect an effect of a given size. Power analysis can also be used to determine power, given an effect size and the number of participants available.

G*Power is a tool commonly used to compute statistical power analyses for *F* tests (Faul, Erdfelder, Buchner, & Lang, 2009). However, it does not allow for a mixed

ANCOVA design to be entered, allowing either a mixed ANOVA or a fixed effect ANCOVA design to be entered. This means it was not possible to calculate power or sample sizes for the recognition task analysis throughout this thesis, where covariates were explored in a mixed design. GLIMMPSE (Muller & Glueck, 2016) was considered as a second option. GLIMMPSE is an open-source online tool for calculating power and sample size for general linear multilevel models. This allows power to be calculated for mixed ANCOVAs, but allows for only one covariate to be included. Three covariates were explored throughout the thesis, with more than one being entered into final univariate models in some instances, limiting the usefulness of GLIMMPSE. GLIMMPSE also requires previous data to be included from previous or pilot studies to complete analyses. This includes expected mean and standard deviation scores, base correlation and decay rate for each factor of within-participant variability, and variability due to the covariate at each level of the repeated measures variable. Some 'best guesses' can be made from previous studies, however, design differences mean these are not likely to be accurate. The cumulative effect of this loss of accuracy across multiple parameters is undesirable, leading to largely imprecise estimates of power. Due to the complex design of the experiments throughout this thesis, power analysis was not possible.

Sample size. Sample size was therefore determined based on the number of variables, interactions, and the resulting number of cells for comparison. The aim was to achieve at least 30 data points per cell, which should lead to approximately 80% power (the minimum suggested power for an ordinary study) given a medium to large effect size (Cohen, 1988). Simmons, Nelson, and Simonsohn (2011) suggest a smaller minimum sample size, arguing that researchers must both decide a rule for terminating data collection before collection begins, and ensure that at least 20 observations are included within each cell, unless a compelling cost-of-data collection justification can be offered. Fewer observations than this are likely to result in inadequate power. Within practice, this is not always achievable, particularly in the context of a doctoral research programme, where time and financial resources may be limited. Recruiting child participants also presents additional challenges in comparison to recruiting adult participants. Achieved sample sizes are discussed within each experimental chapter and the discussion chapter.

Effect sizes. In addition to considering power and discussing achieved sample sizes throughout, effect sizes have also been reported and interpreted throughout Chapters II-V. Within the experimental chapters, η_p^2 has been reported for *F*-tests, *d* for pairwise comparisons, and Pearson's *r* for covariates. Cohen provided guidance for what should be considered a small, medium, and large effect size for *d* (0.20, 0.50, 0.80), and *r* (.10, .30, .50) (Cohen, 1988). When Cohen addresses effect sizes for factorial ANOVA, his account of the partialling out of variables aligns more closely with the modern day use of the term partial eta squared (η^2), rather than eta squared. This has led many textbooks and researchers to report values of .01, .06, and .14 as guidelines for small, medium, and large values of both η^2 for one way ANOVAs and η_p^2 for multiway analysis (for example, Aron, Aron, & Coups, 2008; Coolican, 2009; Howitt & Cramer, 2010).

However, Cohen himself noted that ''The terms 'small', 'medium' and 'large' are relative, not only to each other, but in the area of behavioural science, or even more particularly to the specific content and research method being employed in any given investigation'' (1988). Tabachnick and Fidell (2012) assert that Cohen's guidelines were originally proposed to apply to experiments and social/ clinical areas of psychology; larger values could be expected for non-experimental research, sociology, and the more physiological aspects of psychology. Morris and Fritz (2013) analysed all of the η_p^2 and *d* effect sizes reported within memory journal articles in 2010, the two most commonly reported effect sizes in memory, to estimate memory-specific guidelines for interpreting effect sizes. The quartile figures they obtained for η_p^2 (.08, .18, and .41) were considerably higher than Cohen's values, whilst they were fairly similar for *d* (0.25, 0.57, and 0.99). Whilst it is acknowledged that memory methodology can vary considerably, the data used to obtain these quartiles were domain specific, and included various memory methodologies and paradigms. They are therefore likely to be a more useful guide for interpreting effect sizes throughout this thesis than Cohen's original values. The guidelines suggested by Morris and Fritz (2013) have therefore been adhered to throughout.[1]

What Do We Already Know About Emotional Memory?

The remainder of this thesis examines preferences for emotional content in memory for typically developing children, using the recognition and recall theory and methodologies outlined above. A summary of the existing research into emotional memory in adults and children is provided below.

Evidence of Negativity Preferences in Non-Clinical Populations

Adult negativity preferences. A meta-analysis of laboratory experiments mostly using emotional stories, pictures, faces, or word lists as stimuli examined negativity and positivity preferences in recognition and recall memory amongst adults (Murphy & Isaacowitz, 2008). A significant, medium sized negativity preference in recall was identified across 86 samples of younger adults (M_{age} = 22 years), and a small effect size across 37 samples of older adults (M_{age} = 71 years). There was also a

---

[1] Where covariate main effect sizes are discussed, their overall effect in the model η_p^2 is used rather than their correlation with the outcome measures (*r*). This is to account for other variance in the model, and because memory specific guidelines for *r* were not available.

significant negativity preference in recognition, although only for younger adults. The recognition effect was smaller than for recall, with a near medium effect size observed. The findings suggest that negativity preferences may decrease with age, within adult samples.

Child negativity preferences. As with clinical populations, there is considerably less research into negativity preferences in memory in children compared to adults. The majority of these studies use a DRM paradigm, utilising word lists as stimuli. All but one identified studies found a preference away from negative stimuli during recognition tasks, and report either a preference away from negative material, or do not report net accuracy in recall tasks. Only two studies report evidence of a negativity preference: (1) a recall study using emotional narratives; and (2) a recognition study using images as stimuli.

First, findings are examined where word lists are used within a DRM paradigm. In two studies, discrimination was significantly lower for negative adjectives (Neshat-Doost, Taghavi, Moradi, Yule, & Dalgleish, 1998) and negative words, compared to neutral words (Moradi et al., 2000). Five subsequent experiments found lower hit rates and higher false alarm rates for emotionally negative than for neutral words in recognition tasks, indicating a preference away from emotionally negative stimuli, rather than towards it (Howe, 2007; Howe et al., 2010, Exp 2, 4, and 5). Although analyses on net accuracy (discrimination) have not been reported across these five experiments, the observed lower hit rates, and higher false alarms for emotionally negative words would have resulted in lower discrimination rates than for neutral words, although significance cannot be assumed.

This pattern was found across 8- and 12-year olds (Howe, 2007) and in 7- and 11- year olds (Howe, 2010, Exp 2). These findings were replicated in Experiment 4 with

5- and 8- year olds, and Experiment 5 with 7- and 11- year olds (Howe, 2010). The designs in Experiment 4 and 5 were more complex, including an immediate and delayed recognition task. A three-way interaction in both experiments showed that for neutral lists, hit rates fell but false alarms did not change significantly across the retention interval, whereas for negative lists, hit rates fell at the same rate as for neutral lists, but false alarms increased across the retention interval. These means suggest that memory performance was worse during the immediate recognition for negative compared to neutral material, and decreased at a faster rate across the retention interval, again suggesting a preference away from negative words. Conversely, Quas et al. (2016) found no significant differences in overall net accuracy (discrimination) on a word recognition task depending on valence.

When examining recall tasks using word lists, Howe (2007) found that hit rates and false alarm rates were lower for emotionally negative than for neutral word lists. This effect was replicated in two subsequent studies (Howe et al., 2010, Exp 4 and 5). It is unclear whether this pattern would have resulted in any significant difference in memory performance, because net accuracy is not reported. This pattern was partially replicated in Experiment 2 (Howe et al., 2010), where lower hit rates but higher false alarm rates were identified for emotionally negative than neutral word lists. This again indicates poorer recall memory for negative words. Likewise, a further two experiments found that fewer negative adjectives than neutral words were recalled, with more negative than neutral intrusions (Neshat-Doost et al., 1998), and a significantly lower proportion of negative than neutral words recalled by 11-17 year olds (Moradi et al., 2000).

The only evidence of preferences towards negative stimuli comes from two studies using non-word list stimuli. When short emotional stories were presented as narrative slides to young children (age 5 – 6 years), a negativity preference in recall was shown (Van Bergen, Wall, & Salmon, 2015). Negative images were also discriminated more accurately than neutral images, indicating a negativity preference (Cordon, Melinder, Goodman, & Edelstein, 2013).

To date, the majority of emotional memory experiments in children have employed recognition tasks using word lists as stimuli. Experiment 1 of this thesis presents film clips as stimuli but included no neutral material for comparison. Experiment 2 compares recall performance for negative images and narratives to neutral images and narratives. It is expected that a preference towards negative items will be shown, such as in other studies using similar stimuli (Cordon et al., 2013; Van Bergen et al., 2015). Experiments 3 and 4 of this thesis further explore negativity preferences in recognition memory using word lists. A preference away from negative words is expected based on previous findings.

Evidence of Positivity Preferences

Adult positivity preferences. Murphy and Isaacwitz (2008) also identified a significant positivity preference in recall, with a medium effect for both younger and older adults. However, no positivity preference was found in recognition in either group. Despite Murphy and Isaacwitz not finding a positivity preference in recognition, the sample was small (k = 11), and a further 4 studies using recognition tasks with DRM word lists have been identified which show a positivity preference, and were not included in the meta-analysis but appeared to meet the inclusion criteria (Adelman & Estes, 2013; Doerksen & Shimamura, 2001; Kousta, Vinson, & Vigliocco, 2009; Leigland, Schulz, & Janowsky, 2004). Two of these were published after the meta-analysis was conducted. A further study was also identified which found that positive images are recognized more accurately than neutral images (Charles, Mather, &

Carstensen, 2003). It is unclear whether the addition of these studies would affect the results of the meta-analysis and bring out an overall effect of a positivity preference in adults.

Child positivity preferences. Fewer studies explore positivity preferences in memory in children, compared to positivity preferences in adults, or negativity preferences in children. Two of the three word list DRM studies identified indicate a preference away from positive words, whilst no preference was observed in the final study. Only one experiment using emotional narratives as stimuli found evidence of a positivity preference.

Two word list studies report lower discrimination for positive adjectives (Neshat-Doost et al., 1998) and positive words (Moradi et al. 2000) compared to neutral words in a DRM paradigm. Neshat-Doost et al. (1998) also report that children recalled significantly fewer positive adjectives, and more positive intrusions than neutral words, whilst Moradi et al. (2000) report that children recalled a significantly lower proportion of positive than neutral words. The findings of these two experiments indicate a preference away from positive stimuli in recognition and recall task. As with negative stimuli, Quas et al. (2016) report no significant difference in net accuracy (discrimination) between positive and neutral word lists in 7-8, and 12-14 year olds. The only experiment to show a positivity preference was Van Bergen et al. (2015) whereby positive narratives (in addition to negative narratives) were recalled significantly more accurately than neutral narratives. Due to the limited number of studies in this area, trends must be interpreted with caution.

The limited research in this area and the mixed results in existing literature therefore warrant further exploration into positivity preferences in memory for typically developing children. Recall performance for positive images and narratives has been compared to neutral stimuli in in Experiment 2 of this thesis. A tentative hypothesis is that the positive portions of narrative will be recalled more accurately than neutral portions, in line with Van Bergen et al. (2015). Positive word lists are compared to neutral words in Experiments 3 and 4. It is expected that discrimination will be lower for the positive lists in line with previous research (Moradi et al., 2000; Neshat-Doost et al., 1998).

Evidence of Positive versus Negative Memory Performance

Positive vs. negative memory performance in adults. Murphy and Isaacowitz (2008) report within their meta-analysis that significant positivity and negativity preferences are shown in younger and older adults in recall tasks. Although the effect size remained consistent across the age ranges for positivity preferences, there was a larger negativity effect (compared to the positivity effect) in younger adults, and a smaller negativity effect in older adults. This difference in the negativity effect between younger and older adults was not significant; significance tests of the differences between positivity and negativity preferences were not reported. These findings suggest positivity and negativity preferences remain consistent across adulthood and they appeared equal in size. When examining effect sizes for recognition, all effects were small and non-significant, except for the negativity preference in younger adults. This was a medium effect, significant at $p < .01$. This suggests that negativity preferences in recognition memory are stronger than positivity preferences. A significant difference between the negativity effect in younger and older adults also indicates that this effect decreases with age.

Positive versus negative memory performance in children. There is mixed evidence regarding the difference between memory performance for positive versus negative material in children. When examining recall for word lists, it appears that performance does not differ depending on valence, whereas negative items are recalled more successfully than positive where the stimulus is richer. When examining recognition performance for word lists, findings suggest either no difference in accuracy, or greater accuracy for positive than negative words.

In very young children, negative memory appears more salient than positive memory. Pre-schoolers recall more details about ‘mean’ than ‘nice’ people (Baltazar, Shutts, & Kinzler, 2012), and children aged 2-2.5 years recall a significantly greater proportion of negative, than either neutral or positive material when talking about their own life experiences (Miller & Sperry, 1988). Within this paper, recall was not compared to observed experiences of the young children, and the majority of conversations were initiated by another person. However, there is nothing reported to suggest that the children in this study experienced a high level of negative events in their lives, and children often brought up the negative aspects of the experience themselves even if they did not start the conversation. This finding is in contrast to the observed positivity preference when adults recalled life events (Thomsen, Olesen, Schnieber, & Tønnesvang, 2014). In children age 5-6 years, negatively valenced stories presented as narrative slides were also recalled significantly better than negatively valenced stories (Van Bergen et al., 2015).

Findings are more mixed for recognition of word lists. Two studies found significantly higher rates of discrimination for positive compared to negative words (Brainerd, Holliday, Reyna, Yang, & Toglia, 2010; Moradi et al., 2000), with this effect increasing with age (Brainerd et al., 2010). A further two studies found no significant differences in net accuracy between recognition of positive and negative words in 7-8 year olds, or 12-14 year olds (Quas et al., 2016), or discrimination between positive and negative words in 10-17 year olds (Neshat-Doost et al., 1998).

Recall and recognition performance for a positive versus a negative film is compared in Experiment 1, and between positive and negative images and narrative sets in Experiment 2 of this thesis. Positively valenced material has infrequently been included when exploring emotional memory in children, with no direct comparisons between images or films within existing literature. Recognition performance for positive and negative word lists are compared in Experiments 3 and 4. It is expected that recall will be greater for the negative than positive film, images and narratives in the recall task, in line with previous findings (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015), whilst the pattern of results for recognition is unclear.

Theories of Memory Preferences

It is clear from the experimental evidence presented earlier in this paper that memory performance for emotional content differs to that of neutral content. Further, performance for positive material differs to that of negative. There is no one model which explains the origins and mechanisms involved in various aspects of memory preferences. Whilst evolutionary approaches attempt to provide explanations of why memory preferences have evolved, cognitive models endeavour to undercover the underlying mechanisms.

In addition, preferences appear to change with age; whilst a preference away from negative and positive words is shown in children in recognition and recall, a preference towards negative materials in recognition and recall, and towards positive materials in recall is shown in adults. Negativity effects are stronger than positivity effects in adults, whereas only negative narrative material is recalled at a higher rate than positive narrative material in children. No difference was found when word lists were recalled, and findings were mixed for recognition memory. Developmental theories therefore aim to identify the emergence and progression of preferences across the lifespan. Approaches can be modular, seeking to explain only one aspect of memory preferences (such as negativity preferences), or take a more holistic view of processing.

Evolutionary Models

In essence, evolutionary models assume that animals show a preference towards stimuli of evolutionary significance. One view is that only emotionally negative, threatening stimuli are privileged because they threaten to disrupt the transport of genes between generations (Tooby & Cosmides, 1990). The *Evolved Module of Fear* builds on activation and evolutionary models (Öhman and Mineka, 2001) and is assumed to mediate an emotional level of fear learning. The authors argue that there is dedicated circuitry in the amygdala that is selectively sensitive to any stimuli associated with evolutionarily threatening stimuli encountered in the past, but can also be updated with experience. Once the module is activated, behaviour is modified accordingly. The evolved fear module is automatically activated, and relatively impenetrable to other modules that is has no direct link to. This module allows for maintenance of its activity over time. Evolutionarily advantageous (positive) stimuli are not considered within this module. This model fits with evidence of a negativity preference in adults (Murphy & Isaacowitz, 2008) but is limited in its power to explain positivity preferences, or why this pattern is not observed in children.

An alternative view is that both emotionally positive and negative stimuli have relevance to survival and reproductive success and would therefore be privileged (Hamann, 2001). The *Evaluative Space Model* (ESM) (Cacioppo, Gardner, & Berntson, 1997; 1999) is built around this premise. Norris, Gollan, Berntson and Cacioppo (2010) present a review of the ESM. The model posits that positive and negative affective input is processed differentially with two asymmetries in the activation functions for positive and negative input: (1) the activation function for positivity has a higher intercept than the activation function for negativity (the ‘positivity offset’), resulting in greater positive than negative affect at low levels of affective input; and (2) the activation function for negativity has a higher gain than the activation function for positivity (the ‘negativity bias’). The relationship between positivity and negativity preferences therefore varies according to stimulus strength; when the stimulus is mildly emotional, a positivity preference is typically exhibited, whereas a negativity preference is usually observed for stronger emotional stimuli (Norris et al., 2011). This effect, illustrated in Figure 4, is known as the positive-negative asymmetry (Lewicka, Czapinski, & Peeters, 1992).

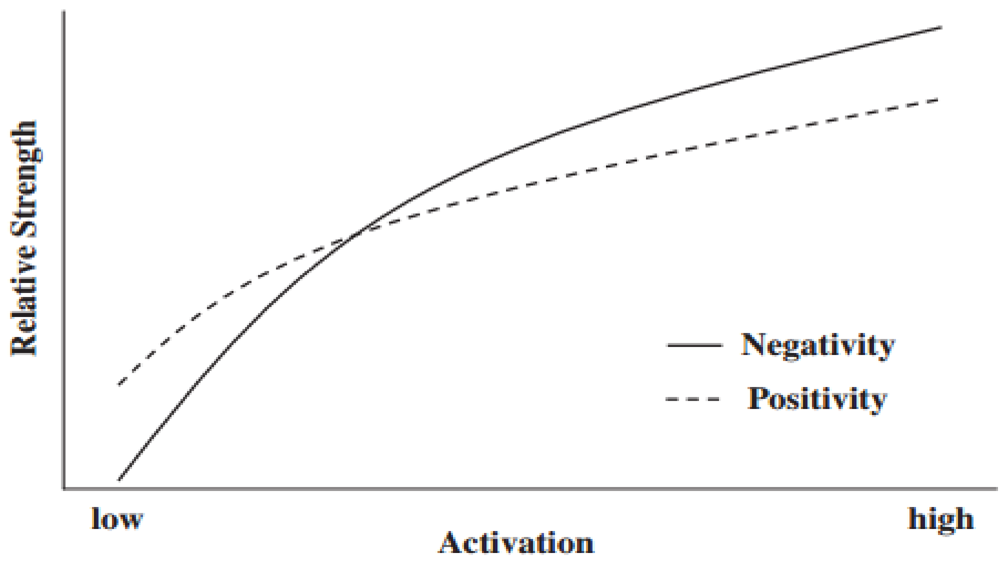


Figure 4. Activation functions for positive and negative dimensions of affective processing; the *x*-axis represents affective input, whereas the *y*-axis represents output of the system. Figure taken from Norris et al. (2011), modified from ‘‘Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis On the Separability of Positive and Negative Substrates,’’(Cacioppo & Berntson, 1994).

This model provides a potential framework for understanding possible behavioural predispositions based on the affective system, and theorizes evolutionary origins for these predispositions, or preferences. A review of available research found converging behavioural and neural evidence of the seperability of positivity and negativity, allowing for multiple modes of activation. Multiple modes of activation allows for co-activation, also supported by a body of behavioural and neural data (Norris et al., 2010). This asymmetry is thought to have evolved as an evolutionary mechanism to promote exploration in the absence of threat, resulting in appetition (approach), which can be observed as a positivity preference, and to maintain vigilance towards threatening stimuli to allow for a quick and targeted response to escape any potentially harmful situations, resulting in avoidance (withdrawal). This can be observed as a negativity preference (Cacioppo & Berntson, 1994; Cacioppo et al., 1997; 1999). As with the Evolved Fear Module, this is only supported by empirical findings of memory preferences in adults. The limited data comparing recognition and recall for positive to negative material appears to support this model, with stronger recognition for positive than negative word lists (Brainerd et al., 2010; Moradi et al., 2000), and stronger recall for negative than positive narratives, where word lists can be thought of as less strong stimuli than narrative materials (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015). There is still limited scope within this model to explain preferences away from positive and negative words in children observed across several studies.

Within the experimental chapters of this thesis, findings of significant negativity but not positivity memory preferences for stimuli would provide support for theories claiming an advantage for only threatening stimuli (for example, see Öhman and Mineka, 2001; Tooby & Cosmides, 1990). Findings of negativity and positivity preferences would indicate that both threatening and emotionally positive stimuli are privileged (for example, see Hamann, 2001). However, where no neutral stimuli are included, these predictions can only be partially tested.

Cognitive Models

Bower's *Network Affect Theory* (Bower & Cohen, 1982; Bower, 1981; Bower, Gilligan, & Monteiro, 1981) postulates that emotions are represented as central nodes in semantic networks within memory. These central nodes have many connections to related ideas, autonomic activity, muscular and expressive patterns and events. Emotional material is encoded propositionally within these semantic networks. Once a central node is activated by emotional material, activation spreads down the prime connections to other related nodes. If activation of these nodes passes threshold levels then they are also activated. Because the central nodes are representations of emotions, the model postulates that emotional stimuli will therefore activate more additional nodes than neutral stimuli. According to this model there should be no difference in memory performance for emotionally negative and positive stimuli, as both are privileged. However, studies have shown significant differences between memory performance for positive compared to negative stimuli in children, in both directions depending on the paradigm used (Baltazar et al., 2012; Brainerd et al., 2010; Miller & Sperry, 1988; Moradi et al., 2000; Van Bergen et al., 2015).

Support for this model can only be partially sought in Experiment 1 of this thesis, whereby no significant difference in memory performance could indicate that positive and negative materials are stored in memory in the same way, but cannot provide support that both are privileged without any neutral stimuli to compare to. Throughout Experiments 2-4, evidence of both a positivity and negativity preference, which are not significantly different to each other, would provide support for this model. However, the underlying mechanism of additional nodes cannot be tested within the scope of these experiments.

Kensinger (2004) provides a more comprehensive review of theories and evidence behind improved emotional memory. Separate processes are hypothesized to account for the facilitation of memory for arousing and valenced materials. During encoding, arousing material is thought to capture attention selectively, evidenced by faster and more accurate reactions to detect threatening, compared to neutral stimuli (Öhman et al, 2001, cited in Kensinger, 2004). Consequently, memory is enhanced for the arousing content, as evident in the ”weapon-focus” effect, whereby memory is enhanced for the threatening weapon, compared to peripheral details (See Buchanan & Adolphs, 2002 for a review). Additionally, arousing material appears to be prioritized during encoding when attentional capacity is limited. This effect was observed in a study which found that arousing words were less likely to be missed during “attentional blinks” (Anderson & Phelps, 2001, cited in Kensinger, 2004). Within memory studies, cognitive load is typically increased by introducing a secondary task to be performed simultaneously. This typically leads to poorer performance in the primary task. However, when arousing materials are used in the primary task, this effect is ameliorated (Bush & Geer, 2001; Kensinger & Corkin, 2004, cited in Kensinger, 2004). Finally, emotional material appears to increase the likelihood of memory consolidation, with emotionally arousing material more likely to be converted to a permanent trace. Memory enhancements for arousing materials therefore tend to be greater after longer delays (Eysenck, 1992; Heuer & Reisberg, 1990; Kleinsmith & Kaplan, 1963; LaBar & Phelps, 1998, cited in Kensinger, 2004). Although this thesis is primarily concerned with valenced stimuli, it is likely that stimuli also varied on arousal, particularly in the more naturalistic experiments (Chapter III) and so the effects of arousal on attention and subsequent memory have been briefly summarised within this review.

Kensinger (2004) notes that in comparison, relatively few studies have explored memory effects for non-arousing, valenced stimuli; but those that have observed a memory boost for valenced material. Kensinger postulates that this is likely the result of conscious encoding strategies, such as semantic elaboration, whereby individuals are more likely to think about items' meanings and their relations to other items, if they are valenced. There may be a number of reasons for increased semantic elaboration; one possibility is that valenced items are typically more semantically cohesive than neutral items. It has been shown that when neutral words and valenced words have similar inter-item associations, the memory advantage for valenced words dissipates (LaBar & Phelps, 1998; Talmi & Moscovitch, in press, cited in Kensinger, 2004).

Developmental Models

Field and Lester (2010) provide an overview of developmental theories of information processing preferences (covering memory, as well as attention and interpretation) towards threatening content in children and adolescents. They propose three main theories of emergence. Firstly, in integral bias models, it is assumed that children are born with varying, innate levels of preferences towards threatening stimuli that remain constant over time. The level of preference is determined by individual factors such as trait anxiety levels. Due to the consistency of preferences over time, this is not strictly speaking, a developmental model. Secondly, moderation models propose that all children are born with high levels of preferences towards threatening stimuli. These preferences are moderated over time by cognitive, social and emotional developmental factors such as anxiety, personality and temperament. This means that they would be moderated to various degrees in different individuals. A review of empirical evidence is provided in (Field and Lester, 2010). Finally, acquisition models are based on the assumption that infants are born without preferences towards threatening stimuli. Instead, these develop further into childhood, or even adulthood, as social, cognitive, and emotional skills develop. The resulting processing preferences are present to a lesser or greater extent depending on an interaction between development and individual factors. Within this model, the level of trait anxiety a child exhibits could either be an individual factor which contributes to the development of processing preferences, or a consequence of these preferences. There is limited research directly assessing acquisition models. However, learning about a stimulus via verbal information increased attentional preferences towards that stimulus in children (Field, 2006a), particularly in anxious individuals (Field, 2006b).

Throughout the experimental chapters in this thesis, age will be considered as a potential covariate in all analyses. Moderation and acquisition models will also be tested by investigating the relationship between discrimination for negatively valenced stimuli and age. For the acquisition model to be supported, the relationship between age and discrimination of negatively valenced stimuli needs to be more positive than the relationship between age and discrimination for neutral stimuli. The reverse pattern would need to be observed to provide support for moderation models. This can therefore only be assessed where a neutral condition is included (the recall task of Experiment 2, and Experiments 3 and 4).

Moderators of Memory

Memory for Different Stimuli Types

Despite only a handful of studies exploring emotional memory in children, some differences begin to emerge when considering the different stimuli sets used. Findings are fairly consistent for DRM studies using word lists. Preferences away from emotionally negative and positive words, towards neutral words are typically observed in both recognition and recall tasks (Howe, 2007; Howe et al., 2010; Moradi et al., 2000; Neshat-Doost et al., 1998). Positive words were discriminated significantly more accurately than negative words in word list experiments, though this finding was not consistent (Brainerd et al., 2010; Moradi et al., 2000).

When more elaborative, narrative stimuli were used, a positivity and negativity preference in recognition was observed (Van Bergen et al., 2015). Recall also appears greater for negative than positive material when more elaborative, narrative stimuli were used (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015). It appears that stimuli type interacts differentially with the effect of valence on memory performance, and that this results in different outcomes depending on type of memory task. It is important to note that these conclusions are based on very small samples and should be interpreted with caution. General memory performance between these stimulus types is explored below, followed by exploration of emotional memory for different stimulus modalities.

Verbal vs. pictorial stimuli. Baddeley proposed in his *working memory* model that there are two different cognitive systems for information processing: a verbal system transmitting and processing sequential information such as written or spoken text, and a visual system responsible for spatial information (Baddeley, 1992). Within recognition and recall memory, pictures are frequently remembered better than words. This is known as the picture superiority effect (Mintzer & Snodgrass, 1999; Paivio & Csapo, 1973).

A variety of DRM studies have been conducted in adults directly comparing memory for verbal/auditory stimuli and visual/pictorial stimuli. Participants recognize and recall images more accurately than words and sentences even when very large pools of stimuli were presented (e.g. Gehring, Toglia, & Kimble, 1976; Shepard, 1967; Standing, 1973). Auditory recognition has also frequently been observed as inferior to visual recognition, for example words and isolated sounds are recalled more accurately when they were presented visually than aurally (Cohen, Horowitz, & Wolfe, 2009; Hilton & Gottwald, 2001). Very little research has been conducted directly exploring differences between auditory and visual memory in children. However, it is well established that younger children rely more heavily on visual working memory in recall, whereas older children rely chiefly upon a phonological component (Hitch, Halliday, Schaafstal, & Schraagen, 1988; Hitch, Woodin, & Baker, 1989). It is therefore expected that children's visual memory would also be superior to verbal memory.

Moving images/film vs. static images. Very few experiments directly compare memory for static images to their moving counterparts. However, several meta-analyses have been conducted comparing the effectiveness of instructional materials presented using different modalities. Those presented through a virtual reality headset were significantly more effective for learning than using computer screens or audio tapes (Ginns, 2005). The sample size for virtual reality headset studies was very small ($n = 4$), so this effect should be interpreted with caution. No differences in learning advantage were found between computer-screen and audio tape tuition material. A further meta-analysis found a medium effect of a learning advantage when instructions were displayed as dynamic rather than static images (Höffler & Leutner, 2007). It is tentatively anticipated that memory for film should be greater than memory for images, which in turn are greater than auditory materials. However, this pattern may differ with emotional material.

Modality and valence. As researchers have begun to move away from simple experimental stimuli, real-life TV clips have been incorporated as complex visual stimuli. A considerable amount of work explores memory for information delivered via TV news broadcasts in comparison to radio or news print. The majority of news broadcasts can be considered emotionally negative, covering topics such as natural disasters, terrorist attacks, and murder. Most earlier work with adult participants showed a superiority effect in recall when news was reported in print, compared to TV or audio presentation (DeFleur, Davenport, Cronin, & DeFleur, 1992; Furnham & Gunter, 1985; Gunter, Furnham, & Gietson, 1984; Wilson, 1974). A minority of studies reported no significant differences between recall of news presented in text or audio-visual format (Stauffer, Frost, & Rybolt, 1980;1981; Wicks & Drew, 1991). It has been proposed that information from print is better recalled because reading is self-paced, and contains less detail, which allows extra time, and encourages the reader to conjure up their own images. These steps result in the stimuli being processed at a deeper level (Craik & Tulving, 1975). By comparison, TV is assumed to be a superficial medium that inhibits deeper processing and detailed recognition of material because of the volume, variability and pace of stimulus presentation (Singer, 1980).

However, more recent studies using child and adolescent participants have found the reverse effect. Children aged 10-11 years recalled more information from a televised news report than from text (Beentjes, Vooijs, & Van Der Voort, 1993), or audio only versions (Gunter, Furnham, & Griffiths, 2000). Fourth and sixth graders recalled significantly more detail from televised news clips than from printed transcripts (Walma Van Der Molen & Van Der Voort, 1997) or news text articles (Walma Van Der Molen & Voort, 1998). However, an earlier study had previously reported the same effect in 4th graders after controlling for reading ability, but found that 6th graders' recall did not differ significantly depending on mode of presentation (Beentjes & van der Voort, 1991). This difference in recall performance for 6th graders is likely due to the content of the clips; Walma Van Der Molen and Van Der Voort (1997) used news reports, whereas Beentjes and van der Voort (1991) used stories. The stories contained emotionally negative content such as a boy falling into a ravine, but were not in the style of a news report. It was therefore presented as fictional rather than factual content. The effect of presentation mode is also ameliorated when children were told before presentation that they needed to learn it for a test (Beentjes et al., 1993).

Fewer studies explore memory for neutral or emotionally positive visual and verbal TV content in children. An early study compared memory and comprehension of audio only, visual only, matched audio-visual, and mismatched audio-visual clips from Sesame Street (Pezdek & Stevens, 1984). Discrimination was higher for the visual only, than audio only conditions, although significance testing was not reported. When presented simultaneously, discrimination was similar for audio and visual elements in the matched condition, but lower for the audio elements in the mismatched condition. Visual discrimination scores remained stable across all conditions regardless of whether it was presented alone or with audio; whereas audio discrimination was significantly lower in the mismatched condition than the matched and audio only conditions.

Summary of modality effects. Research comparing verbal, pictorial and moving image stimuli shows that modality of stimulus, complexity of stimulus, and age interact to affect memory performance. Overall, simple visual stimuli (such as images) are remembered better than simple verbal stimuli (such as individual words), and motion pictures are remembered better than their still image counterparts. However, for more complex visual stimuli such as motion pictures (TV), age moderates the modality effect on memory. Children remember motion pictures better than complex verbal counterparts (narrative or newspapers), whilst this effect is reversed in adults. This fits with the theory that children initially rely more heavily on visual working memory, with this changing to phonological working memory through development (Hitch et al., 1988, 1989).

Experiment 1 in this thesis incorporates positive and negative TV clips, with the memory tasks broken down to compare performance for the visual and narrative components. It is expected that visual elements will be remembered more accurately, but it is unclear how this will interact with valence. Experiment 2 in this thesis incorporates images and audio narratives to further explore this hypothesis with more controlled stimuli, and an added variable of matched or mismatched valence.

Gender

In 1974, Maccoby and Jacklin published a review of psychological gender differences. They concluded that there were no gender differences in learning and memory between boys and girls. However, when looking specifically at episodic memory, 17 of 64 experiments reviewed showed a memory advantage in females, whilst only 2 showed an advantage in males (Herlitz, Nilsson, & Bäckman, 1997). In addition, a large meta-analysis of 165 studies showed a reliable female superiority for verbal ability in girls aged 5 years and under, age 6-10 years, and age 11-18 years (Hyde & Linn, 1988). Longitudinal data in children has shown a relationship between memory capacity and specific verbal skills between kindergarten and second grade, indicating that a verbal skills advantage may also predict a memory advantage (Näslund & Schneider, 1991). Gender has therefore been included as a covariate in all analyses throughout this thesis.

Age

A review of theoretical and empirical research on the development of short term memory across children concludes that there are two distinct short term memory systems; phonological and visuospatial memory (Gathercole, 1998). Phonological short-term memory abilities develop rapidly through the early and middle childhood years. Much of this development appears to arise from developmental increases in the speed of rehearsing and of retrieving material from memory. The capacity to retain information about the visuospatial characteristics of events for short periods of time is mediated by a short-term memory system dissociated from the phonological loop, and may consist of dissociable visual and spatial-temporal sub components. Although young children can retain both kinds of information in memory, their abilities to do so undergo substantial increases between 5 and 11 years of age, when short-term memory capacity approaches adult levels. Experiments using valenced word lists as stimuli have reliably observed an increase in both hit rates, and true recall, but also false alarms and false recall (Howe, 2007; Howe et al., 2010; Howe, Wimmer, & Blease, 2009). Limited differential effects have been reported according to valence.

Outline of the Thesis

Apart from the introduction and concluding chapter, this thesis contains a meta-analysis and three chapters reporting empirical work. The meta-analysis chapter is based on a paper which has been published in *Human Communication Research* (Pearce & Field, 2015).

The meta-analysis reported in Chapter II explores whether an increase in the viewing of “scary” TV could be a contributing factor to the rising levels of anxiety seen in typically developing children (Twenge, 2000). The analysis utilises a multilevel model approach to quantify the overall internalizing effects (fear, anxiety, sadness, and sleep problems) of scary TV on children, and to identify any significant moderators. Findings of the meta-analysis reported in Chapter II support the general perception that violent or scary TV does indeed have a reliable effect on children’s internalizing emotions; however, the impact is relatively low. The main finding shows that increased viewing of scary TV is one of many plausible reasons for the increase in cohort level anxiety, as suggested by Twenge (2000).

This finding raised questions surrounding the mechanisms involved in creating a fear response towards emotionally negative material. An integrative model proposed by Muris and Field (2008) suggests that attention, interpretation and memory distortions may all play a role in the emergence and maintenance of anxiety. Different patterns of attentional (Dudeney et al., 2015) and interpretation (Barrett, Rapee, Dadds, & Ryan, 1996; Marsh, 1986) preferences for threatening stimuli have been shown in clinically anxious, compared to typically developing children, indicating that different processing mechanisms may be involved. Research into memory preferences in children with anxiety is sparse; a meta-analysis with adults found evidence of an increase in recall for negative items in anxious compared to non-anxious adults (Mitte, 2008). There is also evidence of lower memory performance for positive stimuli in some child populations with specific anxiety disorders compared to non-clinical samples (Daleiden, 1998; Moradi et al., 2000; Vassilopoulos, 2012).

Experiments 1-4 therefore explored valence effects on memory in typically developing children. Measures of trait anxiety were recorded to include as a potential covariate in analysis. This enabled the relationship between anxiety levels and preferences toward or away from valenced material to be explored, in non-clinical samples of children. Because no relationship was found in the first three experiments, trait anxiety was not measured in Experiment 4.

Experiments 1 and 2 in Chapter III explore the role of valence, stimulus modality and stimulus congruency on children's memory for narrative and visual aspects of positively and negatively valenced media. The overarching objectives of these experiments are to firstly identify whether 'scary' media is processed differently to positive media within memory, evident as an increase in memory performance for negative, compared to positive material, and second, to explore whether trait anxiety co-varies significantly with any negativity preference found. As anxiety scores increase, an increase in memory performance for negative, over and above that positive material would provide support for the Muris and Field (2008) model of cognitive distortions.

Experiment 1 uses a TV clip with predominately ‘scary’ scenes as a negatively valenced stimulus set designed to elicit a mild internalizing response, and a non-scary clip with predominantly benign and altruistic scenes as a positively valenced stimulus set for comparison. The clips were taken from the TV series *Merlin*, and were designed to provide a high level of ecological validity within a controlled laboratory setting. Prior to the experimental clips being presented, children viewed an introduction clip with images of the main characters and some background information. After stimulus presentation and a brief distractor task, children completed a free recall task. When they indicated they could not remember any more details, they were provided with several cues of character names. Following this, a DRM recognition task was completed where half the items from each clip related to visual aspects, and half to narrative aspects of the clips. The main manipulation of valence was unsuccessful and no differences in discrimination or true recall between the two clips, or modality were revealed. Rotation order appeared to interact with valence and modality effects.

Experiment 2 extends on the findings from Experiment 1 by using valenced still images from the same clips, but matching the images to a narrative of either congruent or incongruent valence, to explore the role of congruency in emotional processing. This allows for two positive and two negative narratives to be compared, in comparison to one clip of each in Experiment 1. It also allows for tighter control of the stimuli sets, because images of the most appropriate exemplars for each valence condition can be chosen, and the narratives can be written to include a portion of neutral information common to both the positive and negative versions. The same procedure was used as in Experiment 1, to allow for comparison between the two experiments.

Valence was successfully manipulated between the conditions with the positive (druid) images, and the negative (monster) images. However, the interaction with modality and congruency shows that the negative narratives were recognized and recalled more successfully than the positive, across the congruent and incongruent conditions. Further, in the incongruent condition only, children recognized the narrative items more accurately than the visual items, and recalled more neutral than emotional material.

The results of the experiments reported in Chapter III provide tentative support for theories of a negativity preference in children, rather than a preference towards all emotional content. One possible reason for the modality effects is the difference in levels of processing between narrative and visual information, whilst the congruency and valence interaction could be explained in part by cognitive dissonance theory and stress coping behaviours.

Chapters IV and V employ a yes/no DRM paradigm to measure accuracy of recognition memory and response bias towards valenced word lists. Experiment 3 (Chapter IV) explores the role of prior emotional context on memory for neutral, positively valenced and negatively valenced words, to extend upon the congruency effect found in Experiment 2. Children were presented with a neutral or valenced story to elicit an emotional context prior to encoding. Children were then presented with emotionally negative, positive and neutral DRM word lists adapted from, or created by Budson et al. (2006) and from the Cortland/Cornell word lists (Brainerd, Toglia, et al., 2008). Immediate recognition tasks were completed after each list was presented. The results do not support the prediction that memory accuracy would be enhanced for words which matched the prior emotional context, but do show a large effect of negative items being discriminated at a greater rate than positive items, when negative items were presented first. The word lists were not fully controlled on other measures such as arousal, imageability and age of acquisition, and it is possible these differences contributed to the effects observed. Additionally, power was assumed to be extremely low once rotation order was included as a fixed effect, making it likely that other true effects in the data were not borne out.

In Experiment 4 (Chapter V) depth of encoding is manipulated to explore its potential role in the mechanism behind the negativity preferences reported in Chapters III, and discrimination difference between positive and negative words in Chapter IV. The same DRM lists were presented as in Experiment 3. A 2 × 2 design was used; children were either instructed to rate the words on spelling (low elaboration task) or emotion elicited (high elaboration task). Orthographic tasks such as assessing spelling are categorised as shallow due to their non-semantic nature, whereas semantic tasks are thought to processed more deeply (Dixon & von Eye, 1984). Children in the emotionally positive groups completed recognition tasks for emotionally positive and neutral word lists, whereas those in the negative groups completed recognition tasks for emotionally negative and neutral word lists. Contrary to expectations, there were no effects of elaboration task on discrimination, and discrimination was poorer for emotional than neutral words, but only significantly so for negative words.

Chapter II

Children's Internalizing Reactions to Scary TV: A Meta-Analysis

Abstract

Public opinion generally portrays 'scary' TV as having a negative impact on children's emotional wellbeing. Existing literature links viewing to sleep disturbances, fear, worry, anxiety and posttraumatic stress disorder. This meta-analysis reviews studies providing a measure of exposure to scary TV, with a primary outcome of internalizing responses. Studies were included only when the entire sample was aged under 18 years. Computerized searches of *PsycARTICLES/PsycINFO*, *PubMed*, *Web of Knowledge* and *Google Scholar*, along with hand searches of journals, books and reference lists were conducted in January 2012, February 2013 and October 2014. One-hundred and seven key authors in the field were contacted to request any unpublished data. Of 131 studies considered, 26 met the full inclusion criteria. A random effects model yielded a small but consistent effect of scary TV viewing on internalizing responses (r = .18), in line with the effect sizes reported for violent TV viewing on externalizing responses. A multilevel model analysis identified that internalizing responses were greater in experimental than correlational studies, studies reporting specific rather than general outcomes, studies using physiological rather than self-report measures, and in samples where all children were aged under 10 years. Implications of these findings are that exposure to emotionally negative stimuli is linked to internalizing responses in typically developing children, in addition to children with anxiety disorders. Future studies should aim to explore the mechanisms behind differences in emotional, compared with neutral information processing in typically developing children.

Children's Internalizing Reactions to Scary TV: A Meta-Analysis

Levels of childhood anxiety are thought to be at an all-time high. Rates of child anxiety have increased over 1 standard deviation between 1952 and 1993 (Twenge, 2000), with the average American child in the 1980s reporting more anxiety than child psychiatric patients in the 1950's. Childhood fear and anxiety is inextricably linked; fears have been associated with subclinical manifestations of anxiety disorders in 49.0% of children interviewed, whilst 22.8% of children met the criteria for at least one DSM-III-R anxiety disorder (Muris et al., 2000). Children between the ages of 7-17 years reported a mean of 13.60 fears in the United States (mean total level of fear 133.20) and 14.29 fears (mean total level of fear 133.84) in Australia (Ollendick, Yang, King, Dong, & Akande, 1996). Although the number of fears are not reported for British children, a similar mean total fear level was reported (M = 138.33), with a similar pattern of top ten fears (Ollendick & Yule, 1990). Additionally, level of fear was found to be significantly correlated to a high extent with manifest (trait) anxiety (r = .64, p < .001).

Limited research has been conducted on the source of these fears and associated rise in anxiety levels. However, Twenge (2000) identified two social indicators which have changed since the 1950s and are significantly linked to an increase in anxiety: social connectedness and a perception of overall environmental threat. Social connectedness draws on research that anxiety can result from social exclusion or a lack of connection in a society (Baumeister & Tice, 1990), or loss of a "sense of community" (Bronfenbrenner, McClelland, Wethington, Moen, & Ceci, 1996). Overall threat relates to individuals appraising the environment and perceiving it as dangerous (Beck, 1985). For example, there has been an increase in reports of violent crimes (U.S. Bureau of the Consensus, 1925-1998), and an increase in fears about nuclear wars (Diamond & Bachman, 1986; Kramer, Kalick, & Milburn, 1983).

The rise in television (TV) consumption has been cited as a possible factor behind a lack of social connectedness and increase in perceived overall threat. In just a 10 year span from 1999–2009, the average amount of TV sets per household has increased from 3.1 to 3.8. The average amount of TV children watched also increased from 3 hr 47 min, to 4 hr 29 min per day, almost an hour of which was viewed via the internet or mobile phones (Rideout, Foehr, & Roberts, 2010). Fifty-three percent of 8-10 year olds and 73% of 11-14 year olds also reported their parents had no rules about how much TV they could view. (Rideout, et al., 2010).

An overall increase in TV viewing has been linked with a decrease in interaction with family members (Vandewater, Bickham, & Lee, 2006). An increase in viewing violent TV has been linked with a decrease in the amount of time spent with friends (Bickham & Rich, 2006), and an increase in anti-social behaviour (Paik & Comstock, 1994). These outcomes meet the description of a lack of social connectedness (Twenge, 2000). The increase in media coverage of threats has also been blamed for an increased perception of environmental threat (Cohl, 2005; Glassner, 2000). For example, over half (61.8%) of 8-10 year old British children sampled were scared of being bombed or invaded, half (49.5%) were scared of earthquakes, whilst 40.1% were scared of death or dead people (Ollendick, Yule, & Oilier, 1991). These fears were also reported at similar levels across children and adolescents in America, Australia, Nigeria and China (Ollendick, Yang, King, Dong, & Akande, 1996). This sort of content is often included on international news broadcasts, as well as in fictional TV content and films. Increased viewing of local news was also a significant predictor of perceived crime threat (Romer, Jamieson, & Aday, 2003). These specific fears would theoretically feed into an enhanced perception of the world as a threatening place, in line with overall threat perception, the second social indicator of increased anxiety levels (Twenge, 2000).

Many TV programmes have also been heavily criticized for containing violent or otherwise explicit content. It is estimated that children in the US are exposed to 200,000 acts of violence on TV by the time they reach 18 years of age (Senate Committee on the Judiciary, 1999). This could lead to an inflated sense of danger from violence in the world, contributing to children's fears.

However, not all fear inducing material shown on TV will include violence. Despite TV and film guidelines (e.g. BBFC, 2014; FCC, 1997; MPAA, 2015) including information on violent content, there is a notable lack of emphasis on non-violent content which might elicit a fear response in children. For example, news programmes are not covered by FCC ratings (FCC, 1997) and can include features such as natural disasters, war, murders and kidnapping. These types of threat were reported as a source of fear in 7-12 year old children, with 70.6% afraid of being bombed/ invaded, 53.8% of fires, 48.0% of burglars and 42.4% of death (Muris, Merckelbach, Meesters, & Van Lier, 1997). This sort of news coverage, along with the high levels of violence seen on TV could lead to a perception of increased levels of overall environmental threat, the second social indicator cited by Twenge (2000) as a factor in the rising levels of anxiety seen in typically developing children.

Some studies have shown that a large portion of children do indeed experience TV or film related fears. Over 75% of 5-6 year olds report TV induced fear, with over 25% of 5-6 year olds reporting that they experience 'great fear' (Korhonen & Lahikainen, 2008). TV induced fears can persist for months and even years; 55% and 35% of typically developing adolescents aged 11-16 years reported enduring fright responses to TV, and films, respectively, viewed when they were aged 6 years or younger (Cantor & Reilly, 1982). Additionally, almost 20% of typically developing adolescents report having seen something on TV that they "couldn't get out of their mind" (Himmelweit, Oppenheim, & Vince, 1958). In line with Twenge (2000), even more adolescents were reporting enduring fright responses 'sometimes' or 'often', with figures up to 25% by the 1980's (Cantor & Reilly, 1982). More recently there have also been anecdotal reports of obsessive thoughts and nightmares related to TV viewing (Cantor & Oliver, 1996). Studies of this kind suggest that fears induced by emotional content on TV may contribute towards the rising levels of anxiety reported by Twenge (2000). However, the literature surrounding the effects of 'scary' TV on internalizing emotions in children is fairly small and disparate. Many older studies are qualitative case studies, and scary TV is often not the primary focus.

This meta-analysis aims to identify whether emotionally negative (scary) TV causes internalizing responses in children. Anxiety and depressive disorders are the primary internalizing disorders; studies with outcome measures relating to anxiety and depression will therefore be included in the meta-analysis. Specific internalizing outcome measures to include are: 1) general fear/ anxiety, measured with validated or non-validated scales; 2) Posttraumatic Stress Disorder (PTSD) as a specific anxiety disorder, measured at a clinical or subclinical level using validated scales; 3) general sadness/ depression, measured with validated or non-validated scales; 4) sleep problems, not specific to any one internalizing disorder, measured using validated or non-validated scales; 5) physiological measures of arousal such as increased skin conductance and decreased heart rate variability, not specific to any one internalizing disorder; and 6) general (unspecified) internalizing response, where global measures of internalizing outcomes have been reported (for example 'total negative emotional and behaviour reactions', including sleep problems, trouble concentrating and avoidance of specific places, measured in Phillips, Prince, & Schiebelhut (2004)).

All outcome measures except 'general internalizing response' have been categorized as 'specific outcome measures' rather than 'general internalizing response' in moderator analysis.

If viewing scary TV is found to cause internalizing responses in children, this would be a first step in demonstrating that it is a plausible explanation of cohort level increases in anxiety. In addition, it will explore various potential moderators in an attempt to begin unpicking underlying mechanisms, should a significant main effect be found.

Moderator Variables

Ten different variables will be considered as potential moderators of the effect of scary TV on internalizing emotions in children. The first set of variables can be grouped together as *study characteristics.* There is a fair amount of overlap between study method, responder and outcome. Correlational studies tend to include self-report measures, typically by parents, or both parent and child. Experiments would more typically include measures directly from the child. The child's behaviour can be observed directly by the experimenter, and preferably would be based on physiological outcomes, either alone or in combination with self-report measures.

An inflated effect size might be expected for correlational compared to experimental studies due to an underreporting of viewing. British adults report watching an average of less than 20 hours of TV per week (TV Licensing, 2011) whereas official viewing figures were over 27 hours per week throughout January 2015 (BARB, 2015). Parents are often required to keep a viewing diary for their children in correlational studies, which is likely to reduce this effect somewhat. Over 70% of mums reported knowing other mums who they judged were letting their children watch more TV than is recommended. Twenty-three percent reported they may not be completely truthful with other mums about how much TV their own child views (Netmums, 2011). This reflects peer-pressure felt by parents not to let their children watch ‘too much’ TV, and a subsequent tendency to underreport viewing. Additionally, 71% of 8-18 year olds sampled in the US reported having a TV in their bedroom, 36% have a computer and 33% have internet access in their rooms (Rideout et al, 2010). Parents may simply be unaware of how much TV their children watch.

Secondly, in correlational studies the stimulus may be more extreme compared to that used in a laboratory. The majority of the studies included in the meta-analysis were published in the 21st century. Experimental studies in recent years have typically been constrained by ethical policies and procedures within universities, with children being classified as a vulnerable sample. This prevents experimenters from exposing children to anything which may cause them distress. Experimenters would not be able to use TV stimuli which had been rated as unsuitable for the sample age group. Thirty-three percent of 8-10 year olds and almost half (49%) of 11-14 year olds state their parents have no rules about what shows they are allowed to watch on TV (Rideout et al., 2010). Combined with the high proportion that has a TV in their room, this means that children will often have exposure to stimuli considered ‘unsuitable’, either with or without parental consent.

Self-report measures are often considered less accurate than physiological measures due to their susceptibility to measurement error, particularly social desirability. Socially desirable behaviours tend to be over reported, and socially undesirable behaviours tend to be under reported, when compared to neutral behaviours. Social desirability was found to influence results in 43% of questionnaire based health-related studies (van de Mortel, 2008). Additionally, parent reports of their children's internalizing emotions are liable to be even less accurate because they cannot access their children's emotions directly. For example, there is 11% mean difference in scores of child negative emotion when reported by adults or children (ICC = 0.55) (Theunissen et al., 1998). Finally, a difference might be expected between studies reporting general outcomes or specific outcomes. Measuring one specific outcome (such as sleep disorder) disregards many other symptoms of negative affect observed in mood and anxiety disorders. Fewer children might appear to be affected when only one specific outcome is measured, but for those who are affected, the effect may appear amplified if an in-depth scale is used, compared to just one or two items on a general questionnaire.

Individual characteristics should also be considered as potential moderators. Gender has been considered in the past when researching fear. While some studies report no significant gender effects on number of reported fears in children (Derevensky, 1974; Eme & Schmidt, 1978), girls report more fears than boys in others (Muris et al., 1997; Ollendick & King, 1991). However, so few TV studies report statistics broken down by gender that it is not possible to include as a potential moderator. Age is another important individual characteristic to consider. Acquisition models assume that as children develop, they start to direct increased attention towards threatening compared to neutral stimuli, and develop an increased tendency to interpret ambiguous information with a threatening explanation rather than a neutral explanation. This model proposes that these behaviours are acquired with experience over a lifetime. Moderation models assume that infants are instead born directing a higher level of attention to threatening compared with neutral stimuli, and favouring threatening over neutral interpretations of ambiguous information. These behaviours can be moderated by experience over time (Field & Lester, 2010). Internalizing responses to scary TV could therefore either increase or decrease with age, respectively.

Finally, *media characteristics* need to be considered. Study age is included as a potential moderator. As previously mentioned, the type and quantity of media consumption is changing over time. Ethical policy for researchers is also changing, limiting the materials which can be used in experiments in more recent years. The type of media could also affect responses; studies have found differences in memory performance for still images compared to words (Shepard, 1967; Standing, 1973), and for moving images compared to still images (Buratto, Matthews, & Lamberts, 2009; Matthews, Benjamin, & Osborne, 2007). This signifies that radio, newspapers and TV are likely to be processed differently. Aspects of media content are also important to consider. Younger children tend to become scared after viewing more magical, mythical, or fictional content with a greater visual component, whereas this shifts to a fear of more realistic events with age (Harrison & Cantor, 1999; Valkenburg, Cantor, & Peeters, 2000; Wilson, 2008). This is hypothesized to be caused by a shift from the preoperational to concrete operational cognitive stages of processing, where children learn to tell the difference between fantasy and fiction more readily (Piaget & Cook, 1952). Finally, a great emphasis has been placed on the effects of violence in TV content, in existing research. It is important to compare studies of scary TV with and without a violent component, to explore whether any internalizing effects could actually be associated with the viewing of violence.

Method

Definitions and Inclusion Criteria

The review was restricted to studies measuring children's internalizing behaviours after viewing 'scary' TV. A study was included only where it met the following criteria:

1. The upper limit of the age range of participants was set to 18 years or under. Any studies reporting an age range with participants over 18 years of age were excluded;
2. Primary outcomes included measures of internalizing responses such as fear, anxiety, worry, sadness and depression. Studies which focused specifically on externalizing behaviours such as aggression and violence were excluded because internalizing emotions were the main focus of the study;
3. Correlational and experimental studies were included in the meta-analysis. Initially only studies using physiological measures or validated scales such as the Beck Depression Inventory (BDI-II) (Beck, Brown & Steer, 1996) were considered for inclusion to obtain high reliability. This criterion was later broadened to include non-validated measures such as parent and child self-report via Likert or visual-analogue scales, due to a lack of studies meeting the stricter inclusion criteria;
4. All experimental studies reported a control group or condition; either a baseline measurement before exposure to scary TV, or a group with no, or limited exposure to scary TV; and
5. Sufficient information was needed from each study to be able to compute effect sizes.

Search Methods

The search strategy is shown in Table 1. An initial search was run in January 2012. Several electronic databases were searched: *PsycARTICLES/PsycINFO*, *PubMed*, *Web of Knowledge* and *Google Scholar,* using a combination of the following keywords: *Child*, *TV* and *Scary*, plus each of the following in turn: *Negative*, *Positive*, *Fear*, *Anxiety*, *Depression* and *Sad*. The library catalogue at Sussex University was searched using the same terms as the electronic database to identify books to be hand searched. In addition, books which included chapters on children's emotional reactions to the media in general were hand searched. The reference lists to any relevant papers or book chapters were also hand searched for additional relevant studies. The initial search identified 101 studies in which children's reactions to scary TV were reported, 80 of which did not meet the inclusion criteria.[2] From these searches, 60 key authors were identified and contacted via email requesting any unpublished data or other papers they had worked on and felt may be relevant. Thirty-one authors responded, but had no further data available to add. The remaining authors either did not respond or could not be reached at their listed contact details. A second search in February 2013 was conducted to check for any papers which had been published since the first search in 2012. This was conducted in the same manner and yielded a further five studies for consideration, one of which was excluded because effect sizes could not be computed.

A final round of searches was conducted in October 2014, prior to submitting the work to a journal for publication. *Happy* was included as an extra internalizing emotion search term, and *Television*, *Film* and *Movie* were added as in turn as substitutes for *TV*. This resulted in an additional 25 studies for consideration, only one of which met all inclusion criteria. Authors who did not respond in the January 2012

[2] For the studies that did not report the relevant statistics to compute effect sizes, authors were contacted to request additional information. One did not respond and the remainder no longer had the original data.

were re-contacted, using any additional contact details that could be found, and a further 47 authors were contacted for any unpublished or additional data for consideration. This yielded two additional published papers and five sets of unpublished data, two of which did not meet all the inclusion criteria because was no direct measurement of TV viewing included.

In total, the three rounds of searches yielded 131 studies for consideration, 26 of which met all the inclusion criteria. One-hundred seven authors were contacted, yielding two further studies (Pfefferbaum et al., 1999; Pfefferbaum et al., 2008) for inclusion and five unpublished data sets for consideration, three of which met all inclusion criteria (Davies & Gentile, 2012; Ferguson, San Miguel, & Hartley, 2009; Gentile & Walsh, 2002). A total of 31 studies were included in the analysis.

Table 1.

Summary of search methods for a meta-analysis of studies into the internalizing reactions to scary TV.

Time of Search	Studies Excluded from Search	Author Responses	Studies Retained
January 2012	Total excluded (*K* = 80)		
Studies identified via search (*K* = 101)	Main outcome externalizing (*k* = 38)	No unpublished data (*n* = 31)	*K* = 21
Authors contacted (*N* = 60)	Review articles/ no new data (*k* = 15)	No response (*n* = 19)	
	Qualitative studies (*k* = 12)	Bouncebacks (*n* = 9)	
	No control group (*k* = 10)	Deceased (*n* = 1)	
	Could not compute ESs (*k* = 5)	Additional data for inclusion (*n* = 0)	
February 2013	Total excluded (*K* = 1)		
Studies identified via search (*K* = 5)	Could not compute ES (*k* = 1)		*K* = 4
October 2014	Total excluded (*K* = 24)		
Studies identified via search (*K*= 25)	Main outcome externalizing (*k* = 6)	Published papers (*n* = 2)	*K* = 6
Authors contacted (*N* = 66)	Review articles/no new data (*k* = 5)	Unpublished data (*n* = 5)	
Repeats (*n* = 19)	Qualitative studies (*k* = 4)	No unpublished data (*n* = 29)	
New (*n* = 47)	No control group (*k* = 4)	No response (*n* = 26)	
	Could not compute ESs (*k* = 2)	Bouncebacks (*n* = 3)	
	Did not measure quantity of TV (*k* = 3)	Retired/ unavailable (*n* = 4)	
Totals			
Studies identified in searches (*K* = 131)	Total excluded from searches (*K* = 105)	Data from authors to be included (*K* = 5)	*K* = 31
Authors contacted (*N* = 107)			

Outcome Measurement

The outcome variable was effect on internalizing behaviours. Separate analyses were intended for each validated scale reported. However, studies used a variety of instruments to measure the same outcome variables. For example, 13 studies measured PTSD symptoms; three used the Impact of Events Scale – Revised Scale (IES-R) (Fairbrother et al., 2003; Pfefferbaum et al., 2000; Pfefferbaum et al., 2001), two used the University of California at Los Angeles Posttraumatic Stress Disorder Reactions Index (UCLA PTSD RI) (Fairbrother et al., 2003; Weems, Scott, Banks, & Graham, 2012), and two used the PTSD Diagnostic Interview Schedule for Children (Phillips et al., 2004; Schuster et al., 2001). The following were each used within just one study; DSM IV criteria (Braun-Lewensohn, Celestin-Westreich, Celestin, Verte, & Ponjaert-Kristoffersen, 2009), the Child PTSD Symptom Scale (Lengua, Long, Smith, & Meltzoff, 2005), the Kiddie Schedule for Affective Disorders and Schizophrenia, Epidemiologic Version (KSADS-E) (Otto et al., 2007), the Parent Response of Posttraumatic Symptoms (PROPS) (Saylor, Cowart, Lipovsky, Jackson, & Finch, 2003), and the Clinical Needs Assessment Instrument (Pfefferbaum et al., 1999). The low numbers of studies using the each validated scale meant this could not be used as a moderator in the analysis.

There were six physiological effect sizes, across two studies. In Gilissen, Bakermans-Kranenburg, van IJzendoorn and van der Veer, (2008) and Gilissen, Koolstra, van IJzendoorn, Bakermans-Kranenburg and van der Veer (2007), physiological responses of increased skin conductance level (SCL-reactivity) and decreased heart rate variability (Root Mean of the Squared Successive differences; RMSSD-variability, computed from raw inter-beat-intervals) were reported. The use of these outcome measures to indicate internalizing responses is consistent with a meta-analysis showing that anxiety disorders are significantly associated with reduced heart rate variability (Chalmers, Quintana, Abbott, & Kemp, 2014), and reports that anxiety and fear increase sweat-gland activity, and subsequently, skin conductance level (Davey, 2014).

There were three effect sizes categorized as 'general internalizing responses', from two studies. Braun-Lewensohn et al., (2009) assessed behavioural and emotional problems using the Hebrew version of Achenbach's Youth Self-Report (YSR; Achenbach and Rescorla, 2001) to give a 'Total Problem Score', and assessed Psychological and Psychiatric Problems using the Brief Symptom Inventory (BSI; Derogatis, 1993). This inventory measures nine dimensions of mental health difficulties, which are summarized by a 'Global Severity Index'. Phillips et al. (2004) used items from several existing scales to measure 'negative emotional and behaviour reactions' including sleep problems, trouble concentrating, and avoidance of specific places. PTSD related symptoms were also recorded but were reported separately. Remaining items were combined into a 'negative reaction' score.

Analysis Procedure

For each study, Pearson's *r* was estimated as the effect size of scary TV viewing on the outcome variables; internalizing behaviours. For correlational studies, *r* was taken directly from the articles. Where *r* was not reported directly, it was calculated from descriptive statistics such as mean, standard deviation and frequencies using common formulae (e.g. Borenstein, Hedges, Higgins & Rothstein 2009; Rosenthal, 1991). This was then converted to Zr as a variance stabilizing transformation (Hedges & Olkin, 1985).

In studies where multiple groups of participants, measures or outcomes were measured, multiple effect sizes were produced. This violates the assumption of independence of effect sizes. The overall analysis and publication bias analysis were therefore conducted with effect sizes aggregated within studies. A random-effects meta-analysis (Hedges & Vevea, 1998) using the DerSimonian-Laird estimator of between study variability (DerSimonian & Laird, 1986) was used for three reasons: first, in social science data, effect size variability is the norm (Field, 2005; Hunter & Schmidt, 2000); second, the random effects model reduces to the fixed effect model if effect sizes do not vary, but not vice versa (Field, 2003); and finally, random-effects meta-analyses enable the results to be generalized beyond the studies included in the meta-analysis (Hedges & Vevea, 1998). The result is a weighted average effect size (and confidence interval) where studies with large samples are weighted more heavily than studies based on small samples. Sensitivity analysis based on Vevea and Woods (2005) was conducted to examine any potential effect of publication bias.

For moderator analysis, a multilevel model was fitted with effect sizes (level 1) nested within studies (level 2). The intercept was a random effect (varying across studies) and moderators were treated as fixed effects. Using a multilevel approach enabled the power of moderator analyses to be maximised by including effect sizes within studies connected to different levels of the moderator variable of interest (Cheung, 2014; Field, 2015; Goldstein, Yang, Omar, Turner, & Thompson, 2000; Konstantopoulos, 2011; Raudenbush & Bryk, 2002).

All analyses were conducted using R 3.1.1. (R Core Team, 2014). The overall analysis was conducted using the *rma()* function in the *metafor* package (Viechtbauer, 2010). The moderator analysis multilevel models were fitted with the *rma.mv()* function in the *metafor* package. The publication bias analysis used Vevea and Woods' (2005) scripts for S-plus adapted for R.

Moderator Variables

Ten variables were included in the moderator analysis: (1) method (experimental vs. correlational); (2) measure (physiological vs. self-report); (3) responder (parent, child, or parent and child report of child's emotional state); (4) outcome measure (internalizing response: fear/anxiety, physiological symptoms, PTSD, sadness/ depression, sleep problems, general/ unspecified internalizing response); (5) mean age; (6) whether the whole sample consisted of children under 10 years of age (all under 10 years vs. all under 18 years); (7) age of study; (8) media type (TV only vs. mixed media, such as TV and radio exposure); (9) media content (fact/news vs. fiction/fantasy); (10) violent content (violent, general, or non-violent). Variables 1-4 comprise study characteristics, 5-6 comprise individual characteristics, and 7-10 comprise media characteristics.

In relation to the violent content moderator, studies were classified as violent if they expressly focused on violent context such as terrorist attacks, assassinations and wars, non-violent if they explicitly did not contain any violence, but otherwise were classed as general (i.e. mixed content).

Results

The average effect size for each study was used to calculate the overall population effect size. Figure 1 shows a forest plot (Lewis & Clarke, 2001) with the effect of scary TV on children's internalizing emotions (*Zr*) for each study and its associated 95% confidence interval. The hashed vertical line indicates no effect.

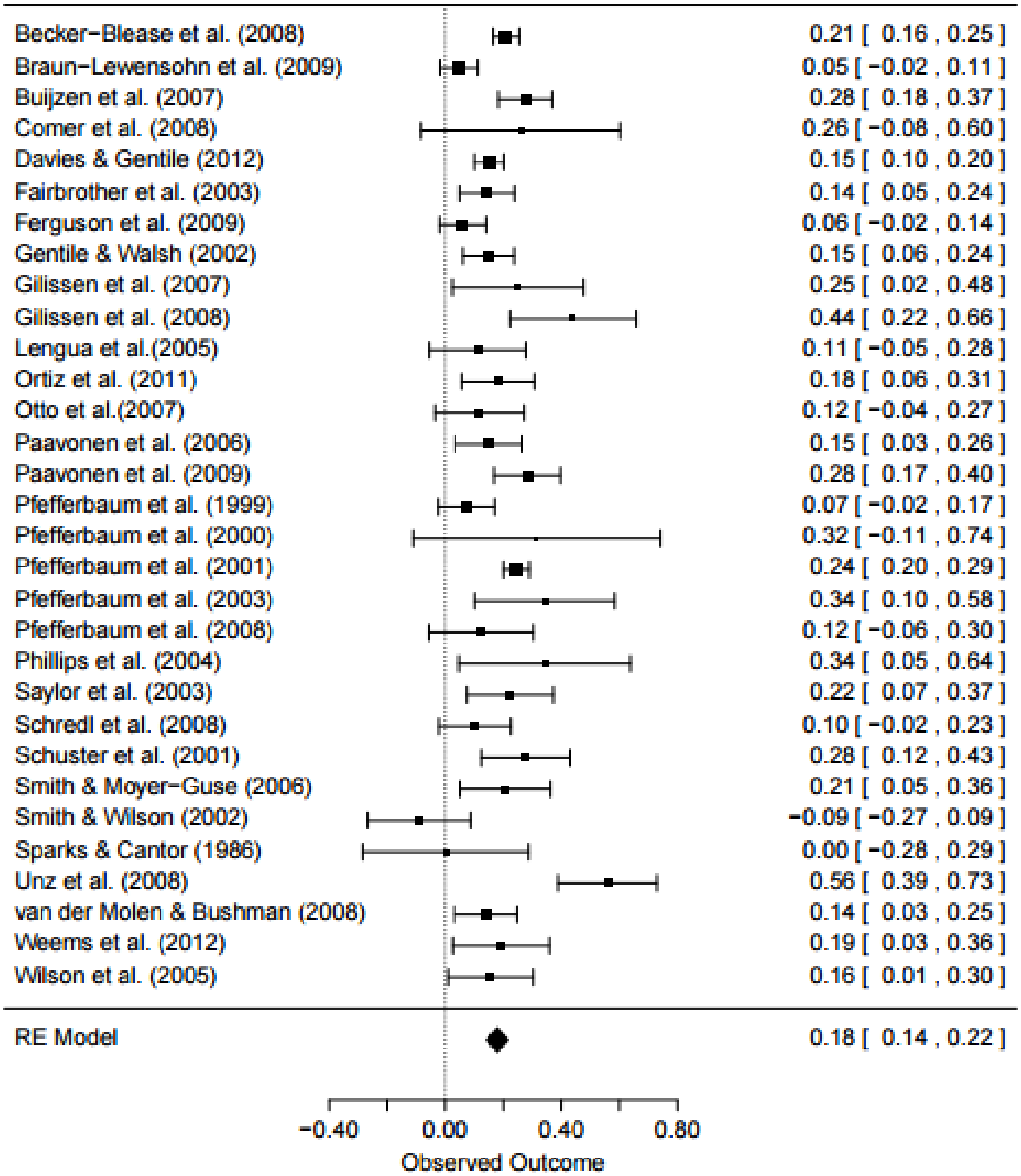


Figure 1: Forest plot of the studies in the meta-analysis. From Pearce and Field (2015).

There was a consistent effect of scary TV on internalizing behaviours, with confidence intervals from 20 of the 31 studies not crossing the hashed line. The overall estimate of the effect of exposure to scary TV on children's internalizing responses is shown as a diamond at the bottom of the forest plot; it shows a modest but consistent effect, $\hat{Z}\rho$ =.18, 95% CI [.14, .22], p < .001. When converted back to r this is an overall effect of r = .18. This would typically be evaluated as a small effect size using Cohen's widely adopted cut-offs, where *r* = .10 (small effect size) would account for approximately 1% of variance and *r* = .30 (medium effect size) would account for approximately 10% variance (Cohen, 1988). There was significant heterogeneity in effect sizes, $Q(30) = 93.38$, $I^2 = 67.87\%$, $p < .001$, indicating variability in study findings that might be explained by moderators.

Publication Bias

Publication bias is a type of selection bias. Publication bias occurs whenever the research that appears in the published literature is systematically unrepresentative of the population of completed studies (Rothstein, Sutton, & Borenstein, 2005). This means a sample would not be representative because the likelihood of studies with certain characteristics being published is higher compared to studies with other characteristics. There are many types of publication bias; this section is concerned with outcome bias. Experiments with positive or significant results are typically more likely to be published than those with negative or non-significant results, resulting in a positive-result bias (Dwan et al., 2008). However, a negative-result bias can also be found in some instances, with papers reporting a negative result significant more likely than papers reporting a positive result to be published in low impact factor journals (Littner, Mimouni, Dollberg, & Mandel, 2005). A lack of publication bias would mean that all studies were equally likely to be published regardless of their outcome.

As noted by Sterne, Becker and Egger (2005), funnel plots are a primary visual tool for the investigation of publication and other bias in meta-analysis, and are appropriate where there are more than approximately five studies included in the analysis. The use of funnel plots necessarily leads to subjective evaluation of selection bias due to visual assessment, however, funnel plots are considered useful in gaining an understanding of the nature of the data. Publication bias causes a distortion of the shape observed on a funnel plot, away from the expected shape. If two tailed selection occurs, the funnel plot will show a sparse area in the neighbourhood of an effect of zero for studies with small sample sizes; both tails may extend beyond that sparse region, so that the funnel appears hollow. When selection is one-tailed, studies with effects near or below zero tend to be missing from the plot, so that the funnel becomes asymmetrical (Vevea, Clements & Hedges, 1993).

Begg and Mazumdar (1994) proposed an adjusted rank correlation method to examine the association between the effect estimates and their sampling variances, and Egger et al. (1997) introduced a linear regression approach in which the standard normal deviation is regressed against its precision to statistically test for funnel plot asymmetry (Stern and Egger, 2005). Begg and Mazumdar (1994) found that their method had only moderate power to detect asymmetry when meta-analyses used a small number of studies (approximately 25). Although Sterne et al (2005) found the linear regression model to be more powerful than the rank correlation method, both were found to have lower power when low number of studies are included and in the absence of severe bias. Lastly, failsafe *N* methods, introduced by Rosenthal (1979) calculates the number of unpublished ‘negative’ studies (studies in which the intervention effect was zero) that would be needed to increase the *p* value for the meta-analysis to above .05. However, all of the above methods work on the assumption that the unpublished or omitted studies, on average, show a null result. Additionally, sample size information is not incorporated into the calculations, widely varying estimates of the number of additional studies are produced when using different methods, and there is a lack of clarity on what constitutes a large *N* (see Becker, 2013 for a review). These methods are therefore not recommended.

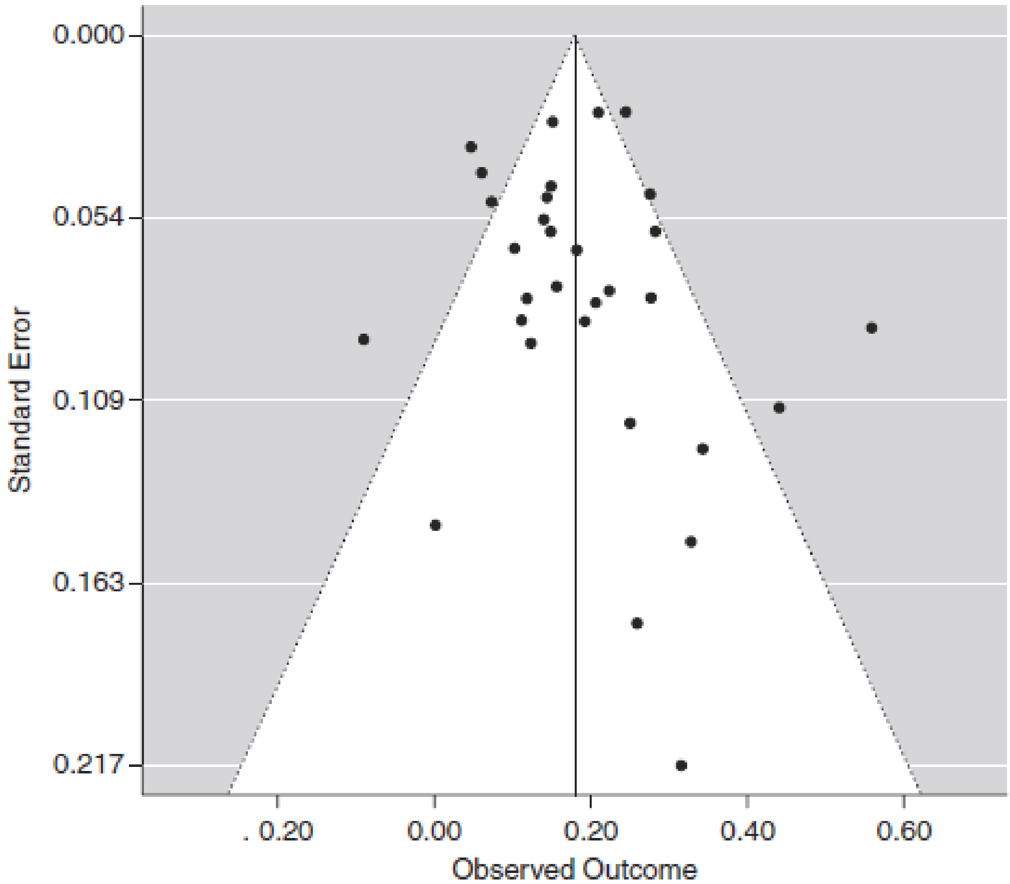


Figure 2. Funnel plot showing effect size of scary TV on internalizing behaviours, plotted against standard error. The vertical line represents the population effect size and the diagonal lines represent the 95% confidence interval. Diagram from Pearce and Field (2015).

Figure 2 shows a funnel plot of the studies in the analysis. In the absence of publication bias, a uniform dispersion of observed outcomes should be shown both above and below the population effect size, within the boundaries of the 95% confidence interval. Funnel plots therefore provide some indication as to whether studies are more or less likely to be published as the observed effect size becomes increasingly smaller or larger in comparison to the population effect size, rather than indicating a bias towards significant or non-significant results being published. This plot indicates likely publication bias, due to the relatively empty space in the bottom left corner; studies with both large *SE*s and effects around and below zero tend to be missing. This deviation from the expected funnel form indicates a 'file-drawer problem': typically, studies with smaller sample sizes produce less precise estimated effects, resulting in larger *SEs* (Sedgwick, 2013). These smaller studies are less likely to be published if they show no significant effect than if they show a significant positive effect (Rosenthal, 1979). In addition, there are several effects falling outside of the funnel (in the grey area). These outliers signify heterogeneity, which could also indicate moderators (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006), alternative selection biases, true heterogeneity, data irregularities such as fraud or inaccurate analysis, artefact, or chance (Egger, Davey Smith, Schneider, & Minder, 1997).

The fact that the bottom left and right panels are very sparse suggests potential one-tailed selection bias, but also studies with small *Ns* (and hence large *SE*s) tend not to show effect sizes at or below zero, which could suggest two-tailed selection bias. Therefore, to quantify the likely impact of publication bias on the population effect size, a sensitivity analysis based on Vevea and Woods (2005) was conducted. Vevea and Woods' sensitivity analysis adjusts the population effect size estimate by applying a sequence of four hypothetical weight functions representing different degrees and forms of possible selection bias. Moderate and severe one- and two-tailed selection bias weight functions were applied because inspection of the funnel plot implied both one- and two-tailed publication.

The original analysis yielded a population estimate of $\hat{Z}\rho = .18$, and the most different adjusted estimate was for severe one-tailed selection bias, $\hat{Z}\rho = .15$. The application of a severe one-tailed bias function represents a pattern of selection that strongly favours the publication of significant positive effects, in which only a small fraction of large negative correlations survive and fewer than half of any correlations near zero are observed. The resulting adjusted estimate shows what the estimate of the population mean effect would have been if this pattern of publication had been observed (Vevea & Woods, 2005). The difference between the adjusted and unadjusted population effect size estimates was relatively small, showing that the estimate of the population mean effect would not be substantially altered. We can therefore be confident that the estimated population effect size has not been severely inflated by unpublished studies missing from the meta-analysis, and no adjustment was needed to correct for publication bias.

Moderator Analysis

Table 2 summarizes the effect size by study, individual, and media characteristics. There were significant, small to medium, positive effects of scary TV viewing on internalizing emotions for studies using all methodology, and measure types, for all responder types, and both age brackets of children, all types of media and content, and all violence categorizations. Significant, small to medium, positive effects were also found for studies reporting all specific outcomes. Studies reporting general internalization as an outcome measure did not result in a significant effect.

Table 2.

Effect sizes of the relationship between exposure to scary TV and internalizing responses by study, individual, or media characteristic

				95% CI	
Moderator	*k*	$\hat{Z}\rho$	*p*	Lower	Upper
Study Characteristics					
Methodology					
Experimental	17	.30	.001	0.13	0.47
Correlational	112	.17	<.001	0.13	0.20
Measure					
Self-report	123	.18	<.001	0.14	0.22
Physiological	6	.37	<.001	0.24	0.50
Responder					
Parent	31	.16	<.001	0.13	0.20
Child	92	.20	<.001	0.13	0.26
Both	6	.21	<.001	0.19	0.23
Outcome					
General internalization	3	.10	.325	-0.10	0.30
Fear/anxiety	68	.19	<.001	0.13	0.25
Physiological	6	.37	<.001	0.24	0.50
PTSD	32	.17	<.001	0.10	0.24
Sadness/ depression	11	.27	.010	0.06	0.47
Sleep problems	9	.18	<.001	0.10	0.27
Individual Characteristics					
Mean age	129	-.01	.432	-.02	0.01
Age under 10					
Under 10	16	.28	<.001	0.17	0.38
Under 18	113	.18	<.001	0.13	0.22
Media Characteristics					
Study age	129	-0.01	.298	-0.02	0.00
Media type[1]					
TV only	97	.19	<.001	0.14	0.25
Mixed media	20	.16	<.001	0.08	0.24
Media Content					
Factual	99	.20	<.001	0.14	0.25
Fantasy	30	.17	<.001	0.11	0.24
Violent Content					
No violence	38	.14	<.001	0.07	0.21
Violence	59	.20	<.001	0.14	0.26
Unspecified	22	.30	<.001	0.16	0.44

[1]*Note:* 12 effect sizes could not be categorized by media type and were excluded from moderator analysis.

Table 3.

Summary of multilevel models showing moderator effects on the relationship between exposure to scary TV and internalizing responses.

Moderator	*k*	Estimate	95% CI Lower	95% CI Upper	Test	*p*
Study Characteristics						
Method (experimental vs. correlational)	129	−0.159	−0.269	−0.049	$\chi^2(1) = .08$	.005
Measure (self-report vs. physiological)	129	−0.177	−0.359	0.005	$\chi^2(1) = 3.64$	.056
Responder	129				$\chi^2(2) = 0.12$	.932
Child vs. Parent		−0.010	−0.067	0.047	$Z = -0.34$	.736
Child vs. Both		0.016	−0.215	0.246	$Z = 0.13$	.895
Outcome	129				$\chi^2(5) = 22.39$	<.001
General vs. Fear/ Anxiety		0.213	0.111	0.315	$Z = 4.11$	<.001
General vs. Physiological symptoms		0.356	0.150	0.561	$Z = 3.39$	<.001
General vs. PTSD		0.126	0.050	0.201	$Z = 3.25$	.001
General vs. Sadness/ Depression		0.244	0.132	0.356	$Z = 4.28$	<.001
General vs. Sleep problems		0.208	0.061	0.355	$Z = 2.77$	.006
Individual Characteristics						
Mean age	129	−0.006	−0.022	0.009	$\chi^2(1) = 0.62$	.432
All in sample under 10 years old	129	0.123	0.001	0.246	$\chi^2(1) = 3.88$	.049
Media characteristics						
Study age	129	−0.007	−0.017	0.002	$\chi^2(1) = 2.31$	.128
Media type (TV only vs. Mixed media)	117	0.035	−0.076	0.145	$\chi^2(1) = 0.38$	.539
Media content (Fact/ news vs. Fantasy/ fiction)	129	−0.017	−0.067	0.033	$\chi^2(1) = 0.45$	.501
Violent content	129				$\chi^2(1) = 2.50$	.287
No violence vs. General		−0.102	−0.256	0.052	$Z = -1.30$	.193
No violence vs. Violent		−0.034	−0.175	0.108	$Z = -0.47$	.640

The multilevel models in which each moderator predicts effect sizes are summarized in Table 3. In terms of study characteristics, the following moderators were significant: (1) methodology, with effect sizes significantly stronger in experimental designs (*k* = 17), $\hat{Z}\rho$ = .30 [.13, .47], *p*<.001, than correlational designs (*k* = 112), $\hat{Z}\rho$ = .17 [.13, .20], *p*<.001; (2) outcome, compared to general indicators of internalizing (*k* = 3), $\hat{Z}\rho$ = .10 [−.10, .30], *p* = .325, effect sizes were significantly stronger for specific measures such as fear/ anxiety (*k* = 68), $\hat{Z}\rho$ = .19 [.13, .25], *p*<.001, physiological symptoms (*k* = 6), $\hat{Z}\rho$ = .37 [.24, .50], *p*<.001, PTSD, (*k* = 32), $\hat{Z}\rho$ = .17 [.10, .24], *p*<.001, sadness/ depression (*k* = 11), $\hat{Z}\rho$ = .27 [.06, .47], *p* = .010, and sleep disorders (*k* = 9), $\hat{Z}\rho$ = .18 [.10, .27], *p*<.001; and (3) measure, there was borderline significance that effect sizes measured using self-report (*k* = 123), $\hat{Z}\rho$ = .18 [.14, .22], *p*<.001, were weaker than those measured physiologically (*k* = 6), $\hat{Z}\rho$ =.37 [.24, .50], *p*<.001.

For individual characteristics, the mean age of participants did not significantly predict effect sizes, however, effect sizes from studies where all participants were under 10 years old (*k* = 16), $\hat{Z}\rho$ = .28 [.17, .38], *p*<.001, were significantly stronger than those that included children aged 10 years and above (*k* = 113), $\hat{Z}\rho$ = .18 [.13, .22], *p*<.001. For media characteristics, all moderators had a small, non-significant, influence on study effect sizes.

Discussion

Main Findings

The meta-analysis found a small but consistent effect of scary TV on children's internalizing responses overall: as amount of viewing of scary TV increased, so did levels of reported internalizing responses. Publication bias analysis shows that despite smaller studies with smaller effect sizes less likely to be published, this had minimal impact on the overall effect size produced by this meta-analysis.

Findings are in line with classic studies and reports which generally portray media, especially scary and violent media, as having a negative effect on children's emotions (Harrison & Cantor, 1999, 1999; Korhonen & Lahikainen, 2008; Mathai, 1983; Valkenburg et al., 2000). The overall effect size is perhaps smaller than would be expected based on classic papers reporting case studies of children showing severe anxiety (e.g. Mathai, 1983) and previous findings that over a quarter of young children experienced 'great fear' in response to TV content (Korhonen & Lahikainen, 2008). However, the effect size for internalizing emotions (r = .18) does fall between the size of the effect of violent TV on externalizing behaviours reported by Wood et al. (1991) (r = .13), and Paik and Comstock (1994) (r = .31).

The key finding of this meta-analysis is that scary TV has a significant effect on internalizing emotions, including anxiety, in children. As scary TV consumption increases, this is likely to lead to an increase in internalizing responses. This fits in with the social indicators theory put forward by Twenge (2000), whereby an increased perception of threat in the environment contributes towards the increased levels of anxiety seen in typically developing children.

Moderators

The meta-analysis found: (1) a significantly greater effect size in internalizing responses reported by experimental than by correlational studies; (2) significantly greater effect sizes where specific rather than general measures were used as the outcome; (3) a borderline weaker effect size when self-report was used rather than physiological measures; and (4) a significantly greater effect size when all the children were under 10 years of age than when older children were also included.

In terms of study characteristics, a larger effect size for internalizing emotions in experimental compared to correlational studies was found. It was anticipated that parents would be likely to underreport TV viewing in correlational studies, leading to an inflated effect size. However, it is also possible that parents would underestimate negative affect experienced by their children because they do not have direct access to their internalizing emotions (Stokes, Pogge, Wecksell, & Zaccario, 2011; Theunissen et al., 1998), or may underreport children's negative affect for fear of appearing a 'bad-parent'. Additionally, experimental studies typically measure the outcome during, or very soon after the stimuli is presented (e.g. Ortiz et al., 2011; Unz, Schwab, & Winterhoff-Spurk, 2008). In comparison, many correlational studies were conducted from several months up to a year after initial TV exposure (Becker-Blease, Finkelhor, & Turner, 2008; Otto et al., 2007). Negative affect may have faded from memory considerably by this point. On a similar note, there was a borderline stronger effect for physiological than self-report measures. This links in with the stronger effect seen in experimental studies than correlational, where physiological measures are typically taken during stimulus presentation in experiments (Gilissen et al., 2007) and can easily be quantified.

As predicted, a larger effect size was reported for specific compared to general outcomes. Specific outcomes are typically measured using validated scales, with high reliability, for example the Post-Traumatic Stress Index used by Braun-Lewensohn et al. (2009) to measure PTSD, and The State–Trait Anxiety Inventory for Children (STAIC) (Spielberger, 1973) used by Comer et al. (2008) to measure anxiety. When a validated scale is used, all aspects of the outcome are included and parents would be more likely to think in depth about the outcome in question. Studies reporting general outcome measures often included a single, or very few questions, such as asking parents to rate negative affect on a 3- or 5-point Likert scale. This could cause confusion for parents between considering the breadth and the severity of various specific outcomes when weighing up how to assign an overall value to the general outcome.

As predicted, stronger effects were found in studies excluding children over 10 years of age. It is unlikely that this is due to younger children watching ‘scarier’ content, as the content would likely be comparable in terms of age appropriateness. Within experimental studies, it is likely that material would be tailored to ensure age appropriateness, due to ethical concerns. Within correlational studies, it is unlikely that children under 10 years watch TV shows with ‘scarier’ content than those over 10 years of age, due to the ratings system used in the UK and US and the scheduling of TV programmes around the watershed. This means material unsuitable for children should not, in general, be shown before 9pm or after 5.30am, when those under 10 are most likely to see it (Ofcom, 2016). In a survey of 1,876 UK parents, 82% claimed they always know what films and television programmes their children watch (ParentPort, 2012), and 83% of parents of 5 – 15 year olds in the UK report having rules in place for their child’s use of television (Ofcom, 2014). This means that younger children responded with greater levels of internalizing responses than older children, despite a lower likelihood that they are viewing material with sexual content, violence, graphic or distressing imagery and swearing. This result could favour a moderation model of fear over an acquisition model; younger children are more likely to direct attention and memory resources to negative or scary, than to neutral TV content, and interpret ambiguous material in a negative fashion, resulting in greater levels of fear. As children develop, these processes can be moderated by experience, and other factors such as social development. This leads to a reduction in attention towards negative or scary over neutral TV material, and less frequent interpretations of ambiguous TV content as threat stimuli (Field & Lester, 2010). It was expected that internalizing emotions would decrease as mean age increased, but this effect was not borne out. Empirical evidence shows a cognitive shift around 7-10 years of age (Cantor & Sparks, 1984; Piaget & Cook, 1952; Valkenburg et al., 2000). The majority of moderation for the processing mechanisms for threat-related stimuli might occur earlier in development, prior to age 10 years, or it could reflect the different materials that younger and older children find scary.

Finally, no significant differences in effect sizes were observed depending on study age, media type, media content (fact or fiction), or whether the stimuli contained violence. It was predicted that older studies would produce greater effect sizes because of the more challenging or unregulated stimuli included. However, newer studies may have been designed with increased methodological rigour, eliminating confounding variables. In terms of media type, it is very difficult to separate correlational studies where children only had access to TV or had access to mixed media. Studies were categorised depending on what exposure measures were reported, but some questionnaires might have only included items on amount of TV viewing and discounted any other forms of media. For example, when exploring the effect of a particular news story in the media (such as a city bombing) the questionnaire might have only included an item on number of times or number of hours of TV exposure, but not included similar items for radio or newspaper exposure. Alternatively, a questionnaire might combine these different media types into one item regarding overall amount of media exposure, which makes it impossible to disentangle their individual contributions. The lack of a significant difference in effect sizes between factual and fictional material is not surprising when considering that younger children are more prone to fear of fantasy items, and older children to fear of factual items (Harrison & Cantor, 1999; Piaget & Cook, 1952; Valkenburg et al., 2000). This implies that one type of content is no scarier than the other, but the relative ‘fear factor’ changes across development. The lack of difference in effect size between scary TV that did, and did not contain violence is somewhat surprising because it was predicted that the inclusion of violence would increase internalizing emotions. However, this could simply imply that both scary and violent content have internalizing effects but do not combine to create a larger effect, rather both are processed simply as ‘negative’ material.

It should also be noted that some moderator sub-groups had small sample sizes (for example, general indicators of internalizing (k = 3), physiological symptoms (k = 6), and both parent and child responding (k = 6)). This leads to a lack of power and an increased likelihood of Type II errors (accepting the null hypothesis when it is false) as well as a reduced likelihood that a statistically significant result reflects a true effect. It is suggested that at least five cases are required to be able to estimate a separate variance component for each subgroup (Williams, 2012).

Implications

The overarching finding from this meta-analysis is that viewing scary TV has a small but significant effect on internalizing emotions in children. In addition, scary TV and film with no violent content is just as detrimental to children’s mental health as that containing violence. Within the context of this thesis, this meta-analysis has identified that levels of state anxiety in children are affected by the viewing of scary TV. The increase in viewing of TV, and in particular, scary TV, is one of many plausible reasons for the rise in cohort levels of anxiety in typically developing children, as suggested by Twenge (2000). Existing literature exploring children’s reactions to scary TV indicates that TV induced fear and anxiety can be severe enough to require hospitalization in children in rare cases (Mathai, 1983; Simons & Silveira, 1994). Some of the studies included in the meta-analysis showed that these effects persist over a prolonged duration. For example, Phillips et al. (2004) reported a significant relationship between amount of televised coverage of the 9/11 attacks viewed and negative reactions three months later, Otto et al. (2007) found that amount of news coverage viewing of the 9/11 attacks significantly predicted PTSD symptoms in children aged under 10 years, one year after the attacks, and Becker-Blease et al. (2008) found significant internalizing responses to televised news coverage of several events up to a year after they occurred. There is also evidence of lasting effects from qualitative studies: 55% and 35% of adolescents report enduring fright responses to TV and films viewed in childhood (Cantor & Reilly, 1982). Up to 90% of adults recall an incident from their childhood when they were ‘really frightened’ by a TV program or movie, and 26.1% of these adults still experienced residual fear or avoidance in relation to the event (Cantor, 1998a, 1998b; Hoekstra, Harris, & Helmick, 1999).

Muris and Field (2008) propose that anxiety is linked to cognitive distortions in attention, memory, and interpretation. There is a considerable lack of research into memory effects of emotionally valenced stimuli in children. A limited number of laboratory studies have been conducted utilising emotionally negative and neutral DRM word lists (e.g. Howe, 2007; Howe et al., 2010; Neshat-Doost et al., 1998; Quas et al., 2015) and images (Cordon et al., 2013) and even less so using emotionally positive words (Brainerd et al., 2010; Moradi et al., 2000; Quas et al., 2015). Many existing studies into the effects of negative TV content typically do not contain positive conditions for comparison (Cantor & Reilly, 1982; Cantor & Sparks, 1984; Hoekstra et al., 1999). Future studies should therefore focus on emotional memory processing in typically developing children, using both emotionally negative and positive materials.

The internalizing effect found in the meta-analysis was greater when measured experimentally, using specific outcome measures, and in children under 10 years.

Future studies should be designed with these features in mind; for example by focussing on children aged under 10 years. In some instances it can provide extremely valuable information to study a naturally occurring event, such as a natural disaster that receives widespread news coverage. However, there are a lot of confounding variables to consider; other sources of exposure, family and child viewing styles, coping styles, and even which channel the news is broadcast on. For example, parents with an active mediation style can successfully reduce the amount of fear and worry younger children experience in response to news coverage, whereas restricting viewing can even increase fear and worry (Buijzen, Walma van der Molen, & Sondij, 2007). To control some of these variables, experimental studies could utilise novel TV programmes (to control for other sources of exposure), and ensure that children view alone (to eliminate parental co-viewing effects), but still use the type of stimuli that typically developing children would be exposed to on a regular basis.

Studies in the Meta-analysis

(Becker-Blease et al., 2008; Braun-Lewensohn et al., 2009; Buijzen et al., 2007; Comer et al., 2008; Davies & Gentile, 2012; Fairbrother et al., 2003; Ferguson, San Miguel, & Hartley, 2009; Gentile & Walsh, 2002; Gilissen, Bakermans-Kranenburg, van IJzendoorn, & van der Veer, 2008; Gilissen et al., 2007; Lengua et al., 2005; Ortiz et al., 2011; Paavonen, Pennonen, Roine, Valkonen, & Lahikainen, 2006; Paavonen, Roine, Pennonen, & Lahikainen, 2009; Pfefferbaum et al., 1999, 2000, 2001, 2003, 2008; Phillips et al., 2004; Saylor et al., 2003; Schredl, Anders, Hellriegel, & Rehm, 2008; Schuster et al., 2001; Smith & Moyer-Gusé, 2006; Smith & Wilson, 2002; Sparks & Cantor, 1986; Unz et al., 2008; van der Molen & Bushman, 2008; Weems et al., 2012; Wilson, Martins, & Marske, 2005).

Chapter III

Children's True and False Memory for Media: Valence, Stimulus Modality and Congruency

Abstract

Despite developmental theories of information processing suggesting a negativity preference in memory, previous experiments report lower recall and recognition rates for negative compared to neutral words. The present two experiments sought to compare the effects of more ecologically valid negative, neutral and positive materials on children's memory. Valenced TV clips were presented in Experiment 1; still images from each clip were matched with positive and negative narratives in Experiment 2. The valence manipulation was unsuccessful in Experiment 1; the only significant finding was a higher number of false memories recalled for the negative clip, when it was presented second. Further research on the clips would be needed, to identify which alterative features might be interacting with order effects. The valence manipulation was more successful in Experiment 2; discrimination and recall was greater for negative than positive narrative material, in line with existing findings. However, there was also a modality and congruency interaction effect that appeared to supersede valence effects. Narrative stimuli were recognized more accurately than visual, where they were incongruently matched. This is discussed in relation to cognitive dissonance and an information seeking approach to coping in stressful situations, as one potential explanation. Future studies should seek to improve the methodological design by increasing power, ensure materials are piloted, and consider any other features of the stimuli which should also be held constant. The interaction between congruency and modality would be an interesting are to research further.

Children's True and False Memory for Media: Valence, Stimulus Modality and Congruency

Chapter II presented a meta-analysis which found a small, reliable effect of scary TV on internalizing emotions in children. Despite the effect being small, some individual studies within the meta-analysis showed that these effects can persist for up to a year after initial TV exposure (e.g. Becker-Blease et al., 2008; Otto et al., 2007; Phillips et al., 2004). When duration and strength of effect are considered together, viewing of scary TV may have a much higher impact on internalizing emotions such as fear, anxiety and sadness, particularly on some individuals. Internalizing emotions such as persistent fear, worry, and sleep disturbances are features of clinical anxiety according to the *DSM-V* (APA, 2013). It is therefore possible that sustained negative effects of viewing scary TV material could escalate to clinical levels of anxiety in some individuals. Anxiety has been linked to distortions in attention, memory, and interpretation (Muris & Field, 2008). Whilst attention and interpretation have been relatively well researched, memory has received less attention. This chapter explores the effects of valenced material on memory, where an enhanced effect of negative, compared to positive or neutral media, would indicator that memory preferences could potentially moderate the link between scary TV exposure and internalizing responses.

This chapter is concerned primarily with episodic memory in children, specifically in relation to viewing emotional media. The majority of laboratory experiments exploring emotional memory in children have thus far utilized DRM word lists as stimuli. Negative emotional media is shown to have an impact on children's internalizing emotions, such as anxiety, fear, and depression (Pearce & Field, 2015). TV clips are therefore considered as a suitable, more naturalistic stimulus for inclusion in memory studies requiring the elicitation of emotional reactions. Preferences towards threatening (negative) stimuli in attention, interpretation, and memory have also been implicated in the emergence and maintenance of anxiety disorders (see Hadwin & Field, 2010 for a review). The following two experiments will explore whether increased memory for negative stimuli can be predicted by an increase in trait anxiety, in line with developmental models of fear acquisition and moderation (Field & Lester, 2010).

Experiment 1 seeks to compare recognition and recall memory in typically developing children for emotionally positive and negative TV clips, taking visual and narrative aspects into account. Experiment 2 seeks to further explore the roles of visual and verbal input, and emotional congruency of multiple inputs, by separating images from narratives and presenting these simultaneously. Congruency is explored as a contextual factor by presenting stories and images which are either matched (congruent) or mismatched (incongruent) on valence. Three potential covariates are explored: (1) age, to test developmental models of fear acquisition and to account for differences in memory performance across childhood; (2) gender, to explore whether girls experience a memory advantage, and to account for any advantages between gender; and (3) trait anxiety, to explore the hypothesis that negativity preferences are stronger in more anxious individuals, and to control for any effect found.

Emotional Memory Enhancement

Early studies into emotional memory focused on memory for real life events, such as being a witness to, or involved in a traumatic event. A meta-analysis revealed that in threatening situations which include the presence of a weapon, people are significantly less likely to identify or to give accurate descriptions of the perpetrator than if no weapon was present (Maass & Köhnken, 1989). This '*weapon focus'* effect postulates that attention is narrowed towards the weapon and away from the periphery. Experimental studies have since sought to explore processing mechanisms for more benign emotional material across the general population, finding clear evidence of a memory preference towards both positive and negative emotional stimuli compared with neutral stimuli in adults (see Buchanan & Adolphs, 2002, and Hamann, 2001 for reviews, and Murphy & Isaacowitz, 2008 for a meta-analysis of empirical findings).

Evolutionary models of information processing, including memory, fall into two camps. Models such as the ESM propose that all stimuli relevant to survival is privileged (Cacioppo et al., 1997; Cacioppo et al., 1999). Alternative models, such as the Evolved Module of Fear propose that only threatening stimuli are privileged (Öhman & Mineka, 2001). Findings reported by Murphy and Isaacowitz (2008) provide evidence that all valenced stimuli is privileged, in support of models such as the ESM.

However, the same pattern of preferences is not observed in children, with the majority of DRM word list experiments showing a preference away from emotionally negative and positive stimuli in both recognition and recall tasks (Howe, 2007; Howe et al., 2010, Exp 2, 4, and 5; Moradi et al., 2000; Neshat-Doost et al., 1998). An additional cognitive 'editing function' has been proposed in recall in which children can filter out more semantically related items compared to neutral items, as an extra step (Howe, 2007; Howe et al., 2010). Although both positive and negative emotional materials have stronger semantic relationships between items, compared to neutral items, literature suggests that this additional filtering is specific to emotionally negative lists; children are more reluctant to talk about negative events and more resistant to false negative suggestion (Goodman & Aman, 1990; Goodman, Hirschman, Hepps & Rudy, 1991). There is limited evidence for this hypothesis from experiments that do not use DRM word lists though. Further, a preference towards recalling negative and positive narratives over neutral has been observed (Van Bergen et al., 2015), and a large body of evidence shows an increase in discriminability as semantic relatedness increases, in adults (Dewhurst & Parry, 2000; Ochsner, 2000; Talmi, Luk, et al., 2007a; Talmi & Moscovitch, 2004; Talmi, Schimmack, et al., 2007). Recall performance should also be lower for negative than for positive stimuli if the editing function hypothesis is correct. Most existing studies into emotional memory with children do not compare positive and negative emotional memory directly. However, of the identified experiments which do, negative material is recalled at a higher rate than positive material when narratives are used as stimuli (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015), though no differences were found when word lists were used (Moradi et al., 2000; Neshat-Doost et al., 1998). This is particularly surprising, given narratives are likely to hold stronger semantic meaning than lists of words.

An integrative model proposed by Field and Lester (2010) suggests that children are either born with varying levels of attention, memory, and interpretation preferences towards threatening stimuli which remain constant with time, are born without these preferences and develop them over time, or are born with high levels of preferences towards threatening stimuli which can be moderated over time (the integral bias, acquisition, and moderation models, respectively). Of these three models, the acquisition model best fits with the empirical evidence, showing more reliable negativity preferences in adults than children. These models will be tested by looking at the relationship between age and memory performance for negatively valenced stimuli within this chapter.

This chapter aims to add to the growing knowledge of emotional memory preferences in children by directly comparing recognition for emotionally positive and negative items, as well as recall for emotionally positive, neutral and negative stimuli. In the first experiment, children will view an emotionally positive clip and an emotionally negative clip. If all evolutionarily advantageous material (both pleasant and unpleasant) is privileged, then no valence effects will be found in the recall task (Experiment 1) and recognition task (Experiments 1 and 2). In addition, higher proportions of both positive and negative narrative will be recalled than the neutral portions in Experiment 2. If only threatening material is privileged, then negative material will be recognized and recalled at a greater rate than both positive (Experiments 1 and 2) and neutral material (recall task in Experiment 2).

Alternatively, if the additional editing function proposed for recall is tenable (Howe, 2007; Howe et al., 2010), recall should be greater for neutral than emotional material, with higher recall of positive than negative material. Similarly, when exploring developmental theories of anxiety emergence, if no relationship is found between age and memory performance for negative material, this will support the integral bias model, a positive relationship will support the acquisition model, whereas a negative relationship will support the moderation model. A positive relationship between age and negativity preference in memory is expected, given the empirical evidence.

False Memory and Emotions

Despite the apparent superiority of emotional memories, false or inaccurate emotional memories are concurrently reported. A false memory refers to remembering and reporting an event that never happened, or an event being remembered differently from the way in which it happened, due to memory distortions (Roediger & McDermott, 1995). Memory distortions become increasingly prevalent with repeated recollections over time (Bartlett & Bartlett, 1995), but can also be reported after just a single stimulus exposure and recollection.

The Deese-Roediger-McDermott (DRM) paradigm is frequently used to measure both true and false memory (Deese, 1959; Roediger & McDermott, 1995).

Within the DRM paradigm, a false memory refers to a false alarm, specifically, a participant stating they remember a lure being presented to them (which had not been presented previously). Studies exploring false memories in children using the DRM paradigm consistently report that although hit rates increase with age, false alarm rates also increase with age for both recall and recognition tasks (e.g. Howe, 2007; Howe et al., 2010, Exp 2 and 4; Howe, Wimmer, & Blease, 2009). A pattern of higher false alarm rates for negative items than neutral items, in addition to lower hit rates for negative than neutral items is found across these experiments. Across time, hit rates for negative and neutral words are found to decrease, whilst false alarm rates remain stable for neutral items, and increase for negative items (Howe et al., 2010), suggesting differential processing. It has also been shown that children are able to suppress false alarms during directed forgetting conditions, whereas adults are not (Howe, 2005). Although changes in false memory across age, and differences between rates of false memory for neutral negative items are consistently reported, less research includes measures of false memory for positive material, or for stimuli other than DRM word lists. The following two experiments therefore explore rates of false recall for positive compared to negative TV clips (Experiment 1), and images and audio narratives (Experiment 2).

Salience of Stimuli

The majority of existing research exploring emotional memory in children has utilized the DRM paradigm, and specifically, word lists as stimuli. Word lists are less salient than pictures, because they have less attributes associated with them and are encoded less distinctly. Some studies have found an advantage for emotional memory performance only when items are presented pictorially (e.g., Talmi, Luk, McGarry, & Moscovitch, 2007b). The only experiment identified which found a negativity preference recognition in children used images as a stimuli (Cordon et al., 2013). Both visual and narrative stimuli will be presented in the current studies, to explore whether increasing stimuli salience will affect emotional memory performance in children.

Ecological Validity of Stimuli

DRM lists have also been criticized for lacking ecological validity, with increases in memory performance thought to be due to details of presentation of words or scenes compared to the individual words (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000; Kensinger & Corkin, 2003). The current studies seek to investigate memory for emotional material in children using stimuli with higher ecological validity compared to regular DRM lists. Children spend a large proportion of their waking hours watching TV with 6-11 year olds watching an average of 28 hours per week in the US (McDonough, 2009), and 4-15 year olds watching an average of 15 hours 42 minutes of television per week in the UK in 2010 (Ofcom, 2014). Eighty per cent of viewing among 10-15 year olds is in 'adult' airtime[3], while just over half (52%) of 4-9 year olds' viewing time is in 'adult' airtime, which may contain emotionally challenging material. A meta-analysis, presented in Chapter II, has shown that there is a small, consistent effect of scary TV on children's internalizing behaviours (Pearce & Field, 2015), and this effect can persist for up to a year after initial viewing (Becker-Blease et al., 2008; Otto et al., 2007). Emotionally charged TV clips have therefore been chosen as stimuli in place of DRM lists in the first study. As well as increased ecological validity, TV clips contain both visual and narrative elements, which can be directly compared in the recognition task in an attempt to tease out any differential effects. Images and narratives have been extracted from these clips for the second study to allow neutral material to be inserted.

---

[3] 'Adult' airtime is defined as the main five PSB channels and ITV, Channel 4 and Channel 5 +1 channels, excluding the slots when children's programmes are shown, combined with all other multi-channel channels except for the dedicated children's channels.

Moderators

Because research into emotional memories in children is at such an early stage, most studies have so far not sought to identify any moderators of emotional memory performance between individuals. Age is explored as a moderator because declarative memory performance increases through early childhood into adolescence (Rittle-Johnson & Siegler, 1998). In addition, differences between emotional memory preferences in adults and children have been identified, and it is unclear at what age these changes develop. Including age also allows developmental theories of fear acquisition to be tested. Gender is explored as a second moderator; a trend is highlighted within adult literature whereby several studies show women to have an advantage in episodic memory studies (Herlitz et al., 1997). This difference is hypothesized to arise from women reflecting on associations within negative lists to a greater degree than men (Dewhurst, Anderson, & Knott, 2012). It is unclear from existing literature whether this effect is reliable, and at what age it may emerge. Previous studies have also pointed towards a link between anxiety and negativity preferences in memory in adults (Mitte, 2008), and in specific groups of anxious children, for specific types of memory tasks (Daleiden, 1998; Moradi et al., 2000; Vassilopoulos, 2012). This is explored as a third moderator.

The aims of Experiments 1 and 2 were to: (1) identify any differences between true and false memories for negative and positive materials in children; (2) explore whether any differences in true and false memories for valenced material vary, depending on type of stimulus; and (3) explore whether memory performance for valenced materials is related to a number of factors, including age, gender, trait anxiety, and congruency of valence for two simultaneous inputs (Experiment 2, only).

Experiment 1: Children's True and False Memories of Emotional Fictional TV

Method

Participants

Forty-three parents of children in school years 3-6 returned opt-in consent forms. Forty-one children completed all conditions of this experiment: 22 boys and 19 girls. Ages ranged from 7 years, 8 months to 11 years, 3 months (*M* = 8 years, 11 months, *SD* = 13.88 months). All children spoke English as their first language and none had a formal diagnosis of any behavioural, developmental or educational disorders according to parent report. All had normal or corrected to normal hearing and vision. The group was ethnically, culturally and socio-economically diverse, including pupils from a Community State School, a Catholic State School and a Progressive Independent School in South England. Two schools were average sized, while one was large. One school had a majority of White British pupils, one a majority of pupils from minority ethnic cultures, and the other consisted of pupils from a diverse mix of cultural backgrounds. One school was noted to have a higher than average proportion of pupils eligible for pupil premium. One school was rated as 'good' and one as 'requires improvement' by their latest Ofsted (Office for Standards in Education, Children's Services and Skills) report, and one was rated as 'good' in the latest ISI (Independent Schools Inspectorate) report. Children took part voluntarily and were naïve to the purpose of the experiment. Children were entered into a prize draw on completion of the study.

Materials

The Multidimensional Anxiety Scale for Children – Short Version (MASC-10), (March, Parker, Sullivan, Stallings & Conners, 1997) was used to measure levels of trait anxiety in children across the four domains of Harm Avoidance, Physical Symptoms, Separation/ Panic and Social Anxiety. This 10-item self-report questionnaire utilises a 4-point Likert response scale from 0 (*Never true about me*) to 3 (*Often true about me*). The scale has been validated for use in participants aged 8 years and above. A minority of children were 7 years old at time of testing, however, the MASC has previously been used successfully with children as young as 6 years old (Cartwright-Hatton et al., 2011). In the current sample, the MASC showed moderate internal consistency, $\alpha = .62$ and was deemed reliable (Bland & Altman, 1997).

Two episodes of the PG certified BBC TV programme '*Merlin*' (Hawes, Frainman, Webb, Orme, & Moore, 2009) were used to create two experimental clips, each 6 min 38 s long, and each containing one coherent storyline, which was predominantly positive, or negative in valence. Clips were edited using Windows Moviemaker. Both episodes were from Series 1, originally aired on British TV in 2008. A 'scary' (emotionally negative) clip, referred to as the 'monster' clip was created from Episode 3, "The Mark of Nimueh". In this clip a witch creates a monster, releases it into the town's water supply and some townsfolk drink the water and die. Merlin and the town physician encounter the monster in a cave, and later go back to fight and kill the monster. The emotionally positive ('Druid') clip was created from Episode 8, "The Beginning of the End". In this clip, a druid boy is lost in the town. Merlin helps the boy using magic and the Prince later helps the boy get out of the town and back home to his family safely. An introduction to the main characters and fictional land of Camelot was created in Microsoft Powerpoint. Pictures of each character were displayed with additional information presented as text and a voiceover. A brief introduction for each clip was created, introducing clip specific characters and providing background context. The main introduction lasted 1 min 19 s, the monster introduction lasted 33 s and the druid introduction lasted 35 s.

A distraction task of ‘spot the difference’ was used for children in school years 3-4 (7-9 years old) consisting of six pairs of cartoon images. There were 3 differences between each pair for children to point to. Children in school years 5-6 (9-11 years old) were given a ‘naming game’. They were presented with a printed copy of the alphabet and asked to name one animal starting with each letter of the alphabet. Children could skip to the next picture or letter if requested. Younger and older children received different tasks because younger children were likely to find the naming task too difficult due to lower vocabulary skills; older children may have found the spot the difference task too easy and become bored, or might have finished it before the allocated time. Children spent 60 s on the distraction task.

A free and prompted recall task was completed; children were asked to verbally describe everything they could remember from each clip in turn. Prompts were included because children are shown to recall more items when there is a discussion about materials they have viewed (Conroy & Salmon, 2006). Children were given prompts relating to three main characters from each clip, for example ‘What do you remember about Prince Arthur? What did he say and do?’ Prompts were open ended because this is a feature of high-elaborative discussion styles, which are also shown to increase recall responses compared to closed questions and low-elaborative discussion styles (Conroy & Salmon, 2006). Responses were recorded on a digital audio recorder, placed on the table between the researcher and child.

The recognition task consisted of a 24-item self-report questionnaire for each clip containing a 5-point Likert scale ranging from ‘0’ (*Sure the statement is false*) to ‘4’ (*Sure the statement is true*) with the midpoint indicating '*Don't know/ can’t remember*'. Children circled their responses on the answer sheet. Twelve items per clip related to visual aspects, for example ‘The witch, Nimueh was wearing a red cape’, and

12 related to story narrative or thematic content, for example 'Merlin looked in some books to find out how to kill the Afanc monster'. Half of the visual and half of the narrative items were true, the other half were false.

Children rated emotions evoked by each clip on a 4-item self-report questionnaire using a 5-point Likert Scale from '0' (*Not at all*) to '4' (*Very, very much*). The four emotional responses recorded were: feeling scared, happy, sad and excited. Fear and happiness were two of the internalizing outcomes included in the meta-analysis in Chapter II. Recording children's ratings for these two emotions enabled a manipulation check, to ensure the negatively valenced clip was causing a similar internalizing emotion to that seen in the previous meta-analysis. They were asked how often they had watched *Merlin* before, whether they remembered seeing either episode before, and whether they would like to watch *Merlin* again, using the same 5-point scale.

Design

A mixed design was used. There were two within-participant independent variables in the recognition task, both with two levels: Valence of TV clip (negative, positive) and Modality of information (visual, narrative). Rotation (positive clip first, negative clip first) was also entered as a fixed effect in analysis. The two dependent variables were hit and false alarm rates, used to calculate *d*' and criterion *C* scores. In the recall task there were two within-participant independent variables, each with two levels: Valence of TV clip (negative, positive) and Valence of memory (negative, positive), with Rotation also entered as a fixed effect in analysis. The dependent variables were the number of false memory utterances, and proportion of true factual material recalled. Participants took part in all conditions, each viewing the positive and negative clip, and each completing the recognition and recall tasks. Age, gender and trait anxiety (MASC-10 scores) were examined in exploratory analyses to see whether their inclusion as covariates was warranted. Analyses for all dependent variables consisted of series' of three exploratory ANCOVAs; age, gender and trait anxiety levels, respectively, were added as covariates separately at each step. They were retained in subsequent steps only when they significantly predicted the outcome measure. Non-significant covariate results are reported in supplementary materials because they are not central to the main findings of the experiments.

Procedure

Interviews were conducted individually in a quiet space at the children's school, away from their classroom. Children had the interview procedure outlined to them along with information regarding confidentiality, anonymity and their right to withdraw. They were given a chance to ask any questions they had, and assent was attained. For all self-report measures except the recall tasks, the researcher read the questions out loud and children circled their response on the answer sheet. Children were told they could ask questions if they did not understand any of the items. Children were monitored closely for any signs of distress, at which point the interview would be terminated immediately with the choice of reading an age appropriate book before the child returned to class.

Children were first asked to answer some questions to measure their personality, at which point they completed the MASC-10. Next, they were asked to watch several sets of pictures and video clips, which they would discuss with the researcher a little later. The introduction to *Merlin* was played first, followed by the first clip introduction and corresponding film clip, and then the second clip introduction and corresponding clip. They were presented on a laptop with a 12" screen, using headphones set to a standardized volume. The laptop was placed on a table approximately 12" from participants. Immediately after viewing the final clip, children were asked to play a quick fun game, at which point they were given the distraction task for 60 s. The order of clip presentation was counterbalanced between participants.

Assent was attained for audio recording prior to the recall task; where assent was not attained, the researcher wrote down children's responses. Children were asked to recall out loud everything they could remember from the first clip, followed by any additional details after each prompt. During free recall, and after each prompt, the next prompt was given if the child spontaneously indicated they could not remember any further details, or if this was indicated when the researcher asked for further details after 10 s silence. Lastly, the emotional ratings for the corresponding clip were completed. The recall and emotional rating tasks were then repeated for the second clip.

The recognition task was then completed for the first, followed by the second clip. Finally, children answered several questions relating to frequency of viewing of *Merlin* and their familiarity with the material. Children were thanked for their time and awarded a sticker. They were given the option of viewing the correct answers for the recognition task before returning to class.

Results

Familiarity with *Merlin*

The majority of children (68.29%) had viewed *Merlin* on TV or DVD prior to the experiment, but most had not seen the episodes from which the clips were edited. Only seven children (17.08%) could recall seeing any part of the positive clip, and nine (21.95%) could recall seeing any part of the negative clip before. Boys reported watching *Merlin* more frequently (M = 1.60, SE = 0.29) than girls (M = 1.14, SE = 0.27), and reported liking *Merlin* to a greater extent (M = 3.05, SE = 0.30) than girls (M = 2.62, SE = 0.26). However, neither of these two differences were significant, $t(39)$ =

1.15, p = .258, 95% CI [−0.35, 1.26], d = 0.36, t(39) = 1.08, p = .288, 95% CI [−0.38, 1.24], d = 0.34, respectively.

Trait Anxiety

Children's normed *t*-scores for trait anxiety (MASC-10) ranged from 30.00 to 67.00 (M = 48.05, SE = 1.29), matching the normal distribution for anxiety in the population (March, 1998). There were no significant differences of reported anxiety scores between males (M = 47.95, SE = 1.81) and females (M = 48.16, SE = 1.87), t(39) = −0.08, p = .938, 95% BCa CI [−5.40, 4.61], d = 0.02. Age was not significantly correlated with anxiety scores, r = −.12.

Emotional Ratings

Rotation order had no significant effect on children's self-reported levels of internalizing emotions (fear and sadness scores combined), F(1, 36) = 0.77, p = .387, η_p^2 = .02. There was no significant difference in children's self-reported internalizing emotions after viewing the negative (M = 0.86, SE = 0.14) and positive (M = 0.78, SE = 0.14) clips, F(1, 36) = 0.41, p = .525, η_p^2 = .01, or significant interaction between rotation and clip valence, F(1, 36) = 0.05, p = .832, η_p^2 = .00. Similarly, rotation order had no significant effect on children's self-reported levels of happiness, F(1, 36) = 0.04, p = .849, η_p^2 = .00. There was no significant difference in children's self-reported happiness after viewing the negative (M = 2.18, SE = 0.18) and positive (M = 2.34, SE = 0.16) clips, F(1, 36) = 0.65, p = .426, η_p^2 = .02, or interaction between rotation and clip valence, F(1, 36) = 0.29, p = .585, η_p^2 = .01.

Recognition Task

Data from three participants were removed from the recognition analysis because their non-response rates were more than 2 *SDs* above the mean, and deemed to be too high for inclusion. The non-response rate was significantly higher for the monster clip (*M* = 0.24, *SE* = 0.03) than for the druid clip (*M* = 0.18, *SE* = 0.03), *t*(37) = 2.11, *p* = .042, 95% BCa [0.01, 0.11]. Table 1 shows the mean number of items recognized for each clip by memory type and modality.

Table 1

Mean number of items correctly recognized for each clip by memory type and modality, out of a maximum score of 6.

Stimulus Type	Positive Clip		Negative Clip	
	M	*SE*	*M*	*SE*
True				
Visual	4.29	0.19	4.14	0.19
Narrative	5.22	0.15	4.46	0.25
False				
Visual	4.19	0.22	3.11	0.20
Narrative	4.10	0.26	3.97	0.26

Data from one participant were removed from recognition analysis because one of their *C* scores was more than 3 *SDs* below the mean. A total of *n* = 22 males and *n* = 15 females were included in the recognition analysis.

Sensitivity: *d*' analysis. Type I *d*' prime was calculated as a measure of how accurately participants recognized true and false details from the TV clips presented to them in the recognition task. Scores were calculated separately for visual, and for narrative items, for each clip, using the method described in Chapter I (see section on Recognition Memory Methodology).

There was a high level of discrimination overall, with mean *d*' scores of around 1.0 or above in all conditions (see Figure 1). One sample *t*-tests showed that participants could discriminate at better-than-chance levels in all conditions, because *d*' scores were significantly greater from 0: positive visual *t*(36) = 10.65, *p* = .001, 95% BCa CI [1.18, 1.75], *d* = 3.55, positive narrative *t*(36) = 16.05, *p* = .001, 95% BCa CI [1.76, 2.19], *d* = 5.35, negative visual *t*(36) = 7.56, *p* = .001, 95% BCa CI [0.69, 1.12], *d* = 2.52, and negative narrative *t*(36) = 10.37, *p* = .001, 95% BCa CI [1.25, 1.78], *d* = 3.46.

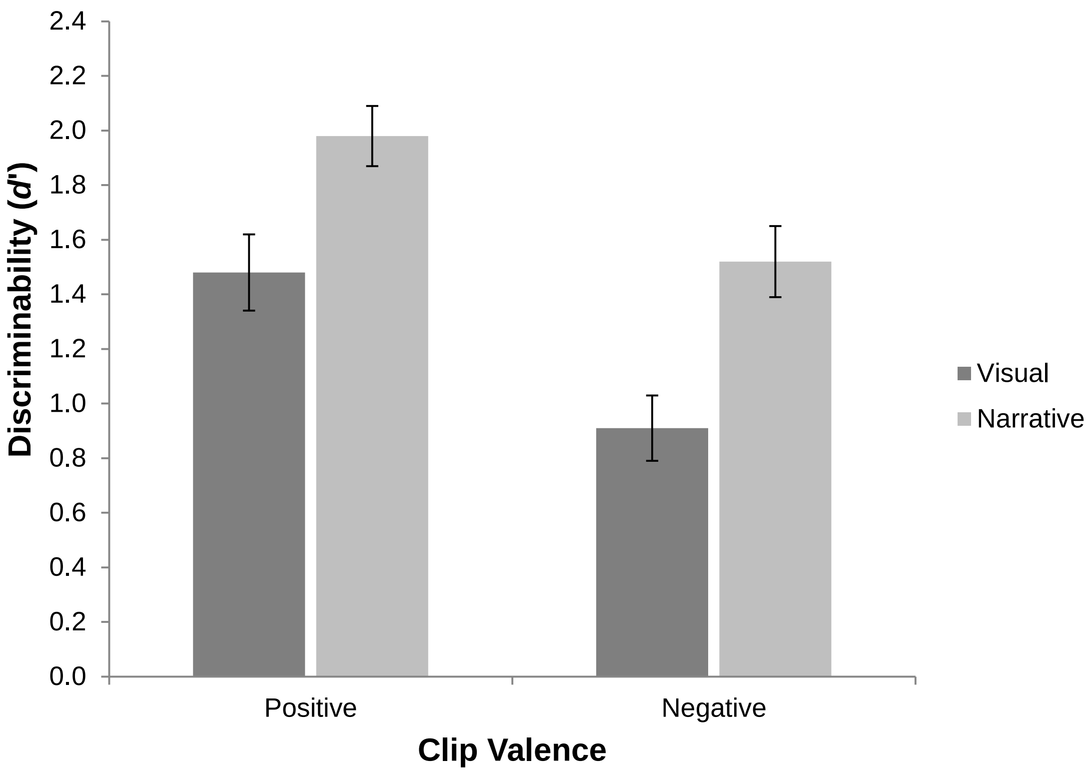


Figure 1. Interaction effects between type of clip and type of modality discriminability rates (*d*' scores). Error bars show standard error.

Age was significantly related to *d*' scores, $F(1, 34) = 6.78$, $p = .014$, $\eta_p^2 = .17$. Discrimination increased with age, $r = .50$. Gender was significantly related to *d*' scores, $F(1, 33) = 11.57$, $p = .002$, $\eta_p^2 = .26$, with boys scoring higher on discrimination ($M = 1.66$, $SE = 0.09$) than girls ($M = 1.17$, $SE = 0.121$). Trait anxiety did not significantly predict discrimination ($p = .469$). [4]

The main analysis was conducted using a mixed 2 × 2 × 2 (Story valence [positive, negative] × Modality [narrative, visual] × Rotation [positive clip first, negative clip first]) ANCOVA, including age and gender as covariates. Whilst average discrimination scores were higher for narrative ($M = 1.75$, $SE = 0.10$) than visual signals ($M = 1.20$, $SE = 0.09$), the main effect of modality on *d*' scores was not significant once

[4] See supplementary materials for full results of covariate analysis.

age and gender had been accounted for, $F(1, 33) = 0.00$, $p = .984$, $\eta_p^2 = .00$. Mean *d*' scores were higher for the positive, druid clip ($M = 1.73$, $SE = 0.09$) than for the negative, monster clip ($M = 1.22$, $SE = 0.08$), but the main effect of valence was not significant, $F(1, 33) = 0.07$, $p = .792$, $\eta_p^2 = .00$. The interaction between valence and modality was also non-significant, $F(1, 33) = 0.28$, $p = .599$, $\eta_p^2 = .01$. The main effect of rotation was not significant, $F(1, 33) = 3.39$, $p = .063$, $\eta_p^2 = .10$, and it did not interact significantly with modality, $F(1, 33) = 0.28$, $p = .599$, $\eta_p^2 = .01$, valence, $F(1, 33) = 0.87$, $p = .357$, $\eta_p^2 = .03$, or modality × valence, $F(1, 33) = 0.04$, $p = .847$, $\eta_p^2 = .00$.

Response bias: Criterion *C* analysis. Response bias has been calculated as described in Chapter I (see section on Recognition Memory Methodology). *C* scores were significantly lower than 0 in the positive narrative condition, ($M = −0.23$, $SE = 0.06$), $t(36) = −3.67$, $p = .002$, 95% BCa CI [−0.35, −0.12], $d = 1.23$, and in the negative visual condition, ($M = −0.22$, $SE = 0.07$), $t(36) = −3.41$, $p = .005$, 95% BCa CI [−0.35, −0.09], $d = 1.14$, indicating liberal response bias. Scores were not significantly different to 0 in the positive visual condition, ($M = 0.01$, $SE = 0.06$), $t(36) = −0.09$, $p = .925$, 95% BCa CI [−0.12, 0.12], $d = 0.03$, or the negative narrative condition, ($M = −0.06$, $SE = 0.08$), $t(36) = −0.85$, $p = .407$, 95% BCa CI [−0.21, 0.09], $d = 0.28$, indicating no response bias. *C* scores are shown in Figure 2.

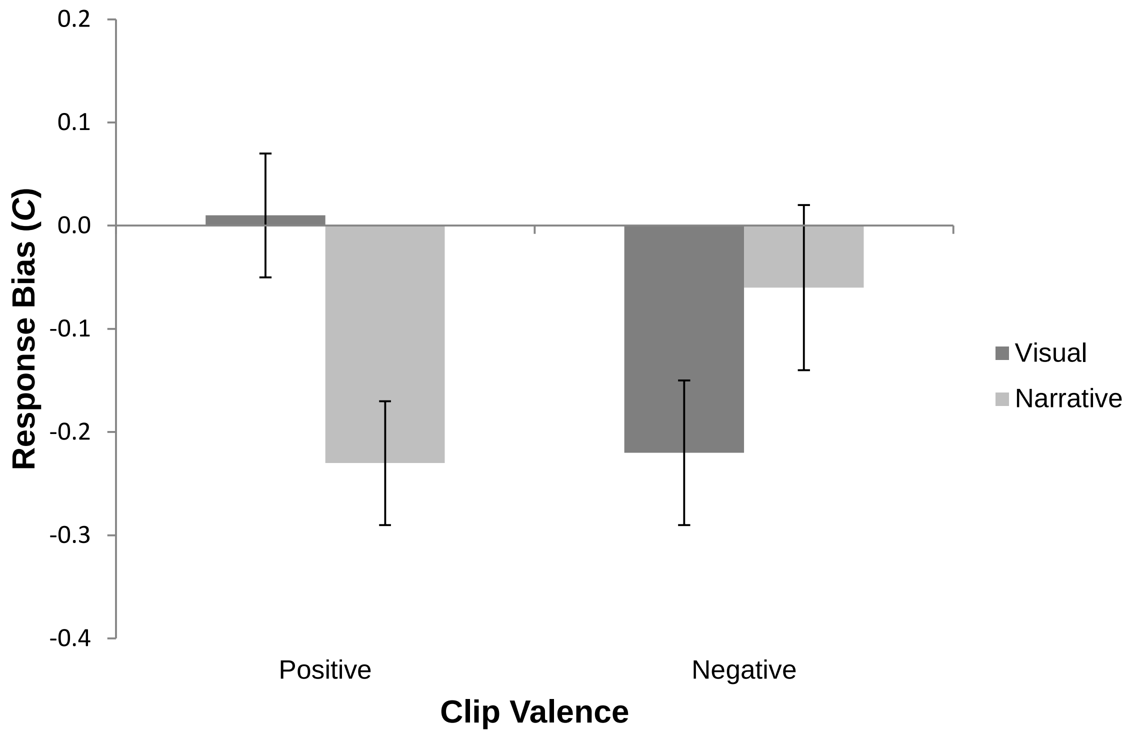


Figure 2. Interaction effects between type of clip and type of modality on response bias rates (*C* scores). Error bars show standard error.

Age, gender and trait anxiety were not significantly related to response bias (*p*s = .999, .924 and .766, respectively).[5] Therefore, the main analysis was simply a 2 × 2 × 2 (Clip valence [positive, negative], Modality [visual, narrative], Rotation [positive clip first, negative clip first]) mixed ANOVA. There was no significant difference in response bias between the positive ($M = -0.11$, $SE = 0.05$) and negative clip ($M = -0.14$, $SE = 0.06$), $F(1, 35) = 0.38$, $p = .543$, $\eta_p^2 = .01$, or between narrative ($M = -0.15$, $SE = 0.06$) and visual items ($M = -0.12$, $SE = 0.06$), $F(1, 35) = 0.41$, $p = .528$, $\eta_p^2 = .01$. The main effect of rotation was not significant, $F(1, 35) = 0.23$, $p = .634$, $\eta_p^2 = .01$, and rotation did not significantly interact with valence, $F(1, 35) = 0.05$, $p = .831$, $\eta_p^2 = .00$, or modality, $F(1, 35) = 1.46$, $p = .235$, $\eta_p^2 = .04$. The valence × modality interaction was significant, $F(1, 35) = 13.44$. $p = .001$, $\eta_p^2 = .28$, but this effect was superseded by a

[5] See supplementary materials for full results of covariate analysis.

significant three-way interaction between valence, modality and rotation effects on *C* scores, $F(1, 35) = 4.87$, $p = .034$, $\eta_p^2 = .12$.

A two-way repeated measures ANOVA showed that for children who viewed the negative clip first there were no significant main effects of valence, $F(1, 17) = 0.37$, $p = .554$, $\eta_p^2 = .02$, or modality $F(1, 17) = 0.14$, $p = .717$, $\eta_p^2 = .01$, on response bias, but the valence by modality interaction was significant, $F(1, 17) = 14.58$, $p = .001$, $\eta_p^2 = .46$. Paired *t*-tests showed significantly greater liberal response bias for the negative visual items than for the positive visual items, $t(17) = 3.57$, $p = .007$, 95% BCa CI [0.16, 0.58], $d = 0.83$, and significantly greater liberal response bias for the positive narrative than negative narrative items, $t(17) = -2.30$, $p = .041$, 95% BCa CI [−0.52, −0.01], $d = 0.59$. For those who viewed the positive clip first, there were no significant main effects of valence, $F(1, 18) = 0.08$, $p = .785$, $\eta_p^2 = .00$, or modality, $F(1, 18) = 2.09$, $p = .165$, $\eta_p^2 = .10$, and the valence by modality interaction was not significant, $F(1, 18) = 1.28$, $p = .274$, $\eta_p^2 = .07$.

Recall task

Children's recall of the clips was transcribed verbatim from audio recordings of their responses. Utterances were chunked and coded according to a pre-defined coding scheme. The narrative of each story was coded into propositional chunks. This acted as a checklist to match each utterance to. Each utterance was categorized in three different ways:

(1) Type of utterance had four possible codes: Factual, emotional, repetition or other. Factual utterances were propositions related to a specific action or utterance by a character. Emotional utterances were those relating to participants' emotions evoked by the clips. Repetition utterances were recorded where children recalled the same factual material more than once. Other

utterances included anything unrelated to the presented clip, or where there was not enough information to otherwise code it accurately. Only factual utterances were further coded for accuracy and valence for subsequent analysis.

(2) Accuracy had two possible codes: True or false. Responses were coded by analysts consulting the detailed written descriptions of the clips and revisiting the clips where necessary.

(3) Valence had two possible codes: Positive or negative. Although the overall clips had been designated as either positive or negative in the recognition task, it was recognized during transcription that there were a minority of positive scenes in the negative clip, and negative scenes in the positive clip. Utterance valence was therefore determined based on the valence of the scene to which the utterance pertained. The coding scheme specified which scenes were designated as positive and negative, and their relevant timings from each clip. The Druid clip contained 287 s (139 chunked items) of positive and 111 s (44 chunked items) of negative scenes. This equates to 75.96% positive, and 24.04% negative items, and a total of 183 chunked items. The Monster clip contained 364 s (106 chunked items) of negative and 34 s (31 chunked items) of positive scenes. This equates to 77.37% negative, and 22.63% positive items, and a total of 137 chunked items.

Reliability of the code scheme was tested by an independent observer coding data for 20% of participants ($N = 8$). There was 91.20% agreement (P_0) between observers for utterance type ($\kappa = 0.58$), 94.87% agreement for accuracy of factual utterances ($\kappa = 0.35$), and 95.26 % agreement for valence of true, factual utterances ($\kappa = 0.91$). P_0 indicates very high agreement between observers for the three sets of coding. The kappa statistic for valence indicates almost perfect reliability. However the kappa statistics indicate borderline weak/ moderate reliability for utterance type, and minimal reliability for accuracy (Landis & Koch, 1977). Low kappa scores are expected when there is a lack of balance across observational categories. It is unsurprising that utterance type and accuracy kappas are low, because the majority of utterances are expected to be factual and true, due to the nature and instruction of the task (Feinstein & Cicchetti, 1995). P_0 is therefore considered a better indicator for reliability. True and false memories were analysed with a one-way (Clip: druid vs. monster) repeated measures ANCOVA; consistent with earlier analyses, age, gender, and trait anxiety were initially entered as covariates, but were retained in the final model only if they significantly predicted memory scores.

False memories. Data from one participant were excluded due to the number of false memories recalled in one condition exceeding 3 *SD*s of the mean. The number of false memories reported increased significantly with age, $F(1, 37) = 4.23$, $p = .047$, $\eta_p^2 = .10$, $r = .34$. Gender and trait anxiety did not significantly predict the number of false memories recalled (*p*s = .461 and .465, respectively).[6] A 2 × 2 (Valence [positive, negative], Rotation [positive clip first, negative clip first]) mixed ANCOVA with age entered as a covariate revealed no significant main effect of rotation order, $F(1, 37) = 2.24$, $p = .143$, $\eta_p^2 = .06$. There was no significant difference between the amount of false memories recalled in the monster ($M = 0.60$, $SE = 0.17$) compared to the druid condition ($M = 0.55$, $SE = 0.13$), $F(1, 37) = 2.77$, $p = .104$, $\eta_p^2 = .07$. However, the clip by rotation interaction was significant, $F(1, 37) = 5.37$, $p = .026$, $\eta_p^2 = .13$.

Paired sample *t*-tests showed for those who saw the druid clip first, they recalled significantly more false memories from the monster (second) clip ($M = 0.62$, $SE = 0.26$), than from the druid clip ($M = 0.14$, $SE = 0.08$), $t(20) = -2.02$, $p = .056$, 95% BCa

6 See supplementary materials for full results of covariate analysis.

CI [−0.95, −0.10], *d* = 0.60. For those who saw the monster clip first, there was no significant difference between number of false memories from the druid (second clip) (*M* = 0.95, *SE* = 0.25) and monster clip (*M* = 0.58, *SE* = 0.23), *t*(18) = 1.10, *p* = .286, 95% BCa CI [−0.33, 1.02], *d* = 0.25. However, significantly more false memories were recalled for the druid clip when it was viewed second, than when it was viewed first, *t*(21.58) = −3.10, *p* = .005, 95% BCa CI [−1.34, −0.27], *d* = 1.01, but false memories of the monster clip did not differ significantly depending on whether they were viewed first or second, *t*(38) = 0.11, *p* = .910, 95% BCa CI [−0.68, 0.75], *d* = 0.04.

True factual memories. The number of utterances correctly recalled from the positive and negative scenes from each story was divided by the total number of positive and negative items in the code scheme to calculate a proportion of material recalled. Data from one participant were excluded due to the proportion of material recalled in the positive druid, and negative druid condition exceeding 3 *SD*s of the mean. On average, the highest proportion of true material was recalled from the negative portions of the monster (mostly negative) clip (*M* = 0.11, *SE* = 0.01). The next highest proportion recalled was from the negative scenes of the druid (mostly positive) clip (*M* = 0.07, *SE* = 0.01). The lowest proportions of material were recalled from the positive scenes of the monster clip (*M* = 0.05, *SE* = 0.01) and positive scenes of the druid clip (*M* = 0.05, *SE* = 0.01).

Age, gender, and trait anxiety scores were not significantly related to the proportion of true material recalled (*p*s = .099, .776, and .324, respectively).[7] However, age did interact significantly with valence of material, *F*(1, 37) = 8.02, *p* = .007, η_p^2 = .18. The proportion of true negative material recalled increased with age at a level approaching significance, *r* = .30, *p* = .058. The proportion of true positive material

7 See supplementary materials for full results of the covariate analysis.

recalled did not change with age, $r = -.06$, $p = .706$. The difference in proportion of positive compared to negative material recalled increased significantly with age ($r = .44$, $p = .006$). Trait anxiety also interacted significantly with valence by clip on the proportion of material recalled, $F(1, 36) = 9.72$, $p = .004$, $\eta_p^2 = .21$. The proportion of true material recalled increased with levels of trait anxiety for positive scenes from the druid clip ($r = .42$, $p = .007$), and for the negative scenes of the monster clip ($r = .30$, $p = .058$). The proportion recalled from the negative scenes in the druid clip ($r = -.05$, $p = .755$) and positive scenes in the monster clip ($r = -.01$, $p = .974$) did not change significantly depending on levels of trait anxiety. Age and trait anxiety were therefore retained as covariates.

A 2 × 2 × 2 (Clip [druid, monster], Valence [positive, negative], Rotation order [druid clip first, monster clip first]) mixed ANCOVA was run with age and trait anxiety scores included as covariates. There were no significant differences in recall depending on whether the positive ($M = 0.06$, $SE = 0.01$), or negative clip ($M = 0.08$, $SE = 0.01$) was shown first, $F(1,36) = 2.67$, $p = .110$, $\eta_p^2 = .07$, between the druid ($M = 0.06$, $SE = 0.01$), and monster clip ($M = 0.08$, $SE = 0.01$) clip, $F(1,36) = 0.45$, $p = .505$, $\eta_p^2 = .01$, or for positive ($M = 0.05$, $SE = 0.01$), compared to negative scenes ($M = 0.09$, $SE = 0.01$), $F(1,36) = 1.68$, $p = .204$, $\eta_p^2 = .04$, after accounting for age and levels of trait anxiety.

The two-way interactions between clip and rotation, $F(1, 36) = 0.20$, $p = .660$, $\eta_p^2 = .01$, valence and rotation, $F(1, 36) = 3.46$, $p = .071$, $\eta_p^2 = .09$, and clip and valence, $F(1, 36) = 1.66$, $p = .207$, $\eta_p^2 = .04$, were all non-significant after accounting for age and trait anxiety levels. The three-way interaction between clip, valence, and rotation on proportion of material recalled was also non-significant, $F(1, 36) = 1.28$, $p = .266$, $\eta_p^2 = .03$, after accounting for age and trait anxiety levels.

Discussion

The stimuli had high ecological validity, compared to many previous emotional memory studies with children. However, this meant it was difficult to ensure comparative proportions of valenced materials were included in each clip; the predominantly negative (monster) clip contained a higher proportion of negative material than the proportion of positive material in the predominantly positive (druid) clip. As a consequence of the clips containing a mixture of valenced material, of uneven proportions; the main manipulation of valence appeared unsuccessful. This meant that observed findings provide only weak evidence to consider, in response to the research aims.

Despite clip valence not being successfully manipulated, some significant differences between the two clips were found in response bias and false recall rates. The clips were designed to be as similar as possible in other key respects; for example, length, TV series, main characters, number of additional characters, and dates originally aired. It is possible that the measures of internalization and happiness were not sensitive enough to identify a difference in valence, but it could also be the case that some other difference between the two clips resulted in the observed differences in memory performance.

This work could be extended by including a neutral control condition to allow emotional memory theories to be tested more fully. The clips were edited directly from TV episodes of *Merlin*, which is written as an adventure TV series. This left little scope to include a neutral control clip, or to edit neutral portions of narrative material into the clips for comparison. Positivity and negativity preferences in memory could not be explored, because there were no baseline measures. Comparisons were therefore limited to relative differences in memory performance between positive and negative stimuli.

Experiment 2: Children's Memories of Emotional Images and Stories

Experiment 2 is intended to improve upon the design of Experiment 1, whilst maintaining the use of naturalistic stimuli. The same two clips from Experiment 1 are used as a basis for the visual and narrative stimuli, but two versions of each story have been written to accompany a set of still images taken from each clip. This has resulted in two positive and two negative stories, and 15 positive and 15 negative images. The aim of this is to ameliorate the effects of any extraneous variables which may not have been otherwise accounted for. This should help to ensure that any observed effects are due to valence, rather than other factors (e.g. salience). The two versions of each story share some neutral narrative content so that memory of both positive and negative content can be compared to that of the neutral content. This provides a baseline measure to explore positivity and negativity preferences. An equal amount of valenced material is also included within each story pair, whereas the amount of positive to negative material in each clip of Experiment 1 was unbalanced.

This methodology is similar to the famous 'slides procedure' utilised by Cahill, Prins, Weber, and McGaugh (1994) to study emotional memory in adults. In Cahill et al's experiment, 12 slides were presented, paired with either a 'neutral' or 'arousing' (negative) narrative. One sentence was presented with each slide. The first four slides and the final slide were paired with identical narratives. The remaining seven slides used similar sentences but varied key phrases between the two versions. Whilst the study was testing the effect of Propranolol on emotional memory, results from the placebo group suggest that recognition scores (percent correct) were higher across all slides in the arousing (negative) compared to the neutral group. Additionally, both groups had the highest recognition for slides 5-9, followed by slides 10-12. Reported results do not indicate whether these effects are statistically significant. It is difficult to draw any hypotheses for the current study based on these findings because the experiment included adults rather than children, compared a negative to neutral story, rather than a negative to positive story, and did not draw a firm distinction between arousal and valence in the negative condition.

Method

Participants

Students from two schools in East Sussex, England participated in the study. Fifty-nine parents of children in school years 3-6 returned opt-in consent forms. Three children did not complete the interview in its entirety and were excluded from all analyses. Thirty-three boys (58.93%) and 23 girls (41.07%) completed the experiment. Ages ranged from 7 years, 1 month to 11 years, 1 month (*M* = 8 years, 11 months, *SD* = 13.79 months). All spoke English as their first language and none had any formal diagnosis of behavioural, developmental or educational disorders based on parental report. All children had normal or corrected to normal hearing and vision. The group was ethnically, culturally and socio-economically diverse including pupils from a mixed Community State School and a mixed voluntary aided Church of England School. One school was average size, whilst the other was larger than average. The majority of pupils at both schools were White British, but one school reported 15 different additional languages spoken throughout the school. Both schools had a lower than average proportion of students eligible for pupil premium and were rated as 'good' by Ofsted. Children took part voluntarily and were naïve to the purpose of the experiment. They were entered into a voucher prize draw on completion.

Materials

Trait anxiety. The MASC-10 (March, Parker, Sullivan, Stallings & Conners, 1997) was used to measure levels of trait anxiety (see Experiment 1). The scale showed poor internal consistency in the current sample, α = .54 (Bland & Altman, 1997). Although lower alphas can result from shorter scale length and a narrower scale of responses, this is shown to have a more noticeable effect for scales with fewer than 7-items, and response scales narrower than 4-points (Cronbach, 1951; Voss, Jr, & Fotopoulos, 2000). The MASC consists of 10 items and a 4-point response scale, so a higher alpha is expected. The MASC was deemed unreliable and scores excluded from analysis.

In addition, the Spence Children's Anxiety Scale - SCAS (Spence, 1997; 1998) was administered. The SCAS is a 45-item scale measuring six domains of anxiety, and including six filler items. Muris et al. (2000) found evidence to support the psychometric properties of the SCAS in 7-11 year olds, indicating its suitability within this experiment. Items were read aloud to children, who were asked to circle how often they experienced each one on a 4-point scale (never = 0, sometimes, often, or always = 3). A total score was calculated by summing responses of the 39 anxiety items, and converting to *t*-scores. The SCAS had good internal consistency in the current sample, α = .84.

Emotional stimuli. The same episodes of *Merlin* (Hawes et al., 2009) were used as in Experiment 1 to create the four sets of experimental stimuli (see Appendix: Experiment 2). An emotionally ‘positive’ set of 15 still images were taken from Episode 8, "The Beginning of the End" and is referred to as the ‘Druid’ story condition throughout, regardless of the paired story valence. Images included some boots standing behind a pillar, some keys hovering in the air, a bowl of soup with keys in them, and a boy being hugged by his parents. Fifteen emotionally ‘negative’ images were taken from Episode 3 “The Mark of Nimueh”, referred to as the ‘Monster’ story condition throughout. It included images of a witch, an animal growing inside a glowing sphere, characters visiting dark caves with swords, and a monster appearing from a water pool. Each image set was made into a Powerpoint presentation, with each image presented full screen on a laptop for 20 s. Original stories from the *Merlin* episodes were adapted to provide an emotionally positive, and emotionally negative version of narrative to be matched with each image set. A substantial portion of the positive and negative versions of each story was neutral and identical, with key phrases changed to alter the valence of the story.

In the positive druid boy story, a boy is in Camelot with his family to pick up supplies and plays a game with the other children to see who can get back home the fastest without being spotted. Merlin and Prince Arthur help the boy get home and he wins for the first time. In the negative version, the druids are in Camelot to kill the King. The boy gets lost and is hiding so he will not be imprisoned or killed. Eventually Prince Arthur helps the boy escape back to his family, where they continue to plot against Camelot.

In the positive monster story, a nice witch creates an animal to live in the caves below Camelot to help her look after the villagers who keep falling asleep in the daytime, after drinking the village water. The animal works out there is a bug in the water making them sleepy, and helps the witch clean the water. In the negative version, the witch creates the monster to poison the water and kill the people in Camelot. Some villagers are found dead, so Merlin and the doctor search for the monster. The Prince tries to kill it, unsuccessfully.

The stories were an average of 837 words long, with a difference of eleven words in length between the positive and negative druid boy story, and a difference of six words between the positive and negative version of the monster story. The stories were audio recorded, and inserted into the image presentations. An average of 4-5 sentences were heard whilst each image was presented.

The same character introduction and distraction tasks from Experiment 1 were used. The same recall and recognition tasks were also used, with some minor wording alterations between positive and negative versions (such as the use of ‘monster’ or ‘animal’) to match the phrasing used in the respective stories. As in Experiment 1, participants rated emotions evoked by each set of images and each audio story, and were asked the same questions regarding their viewing habits and preferences in relation to *Merlin*.

Design

A mixed design was used. There were three independent variables in the recognition task, all with two levels: Story (monster, druid), Congruency of image/ story valence (congruent, incongruent), and Modality of information (visual, narrative).[8] Story and modality were within-participants variables, and congruency was a between-participants variable. In the congruent condition, children viewed the druid (positive) images with the positive version of the druid story, and the monster (negative) images with the negative version of the monster story. The images were paired with the mismatched valenced story in the incongruent condition. The two dependent variables in the recognition task were hit and false alarm rates, used to calculate *d*’ and criterion *C* scores.

Within the recall task, false memories, true visual memories, and true narrative memories were analysed separately. Across all analyses there were two independent variables: Story (monster, druid), and Congruency of image/ story (congruent, incongruent). Story was a within-participant variable and congruency was a between-

[8] Because rotation would be too highly correlated with congruency, it has not been entered as a fixed effect in analysis.

participant variable. For true narrative recall there was a third, within-participant independent variable: Material type (emotional, neutral). The dependent variables were measured by the number of utterances which related to false, or true visual content from the code scheme. For true narrative recall, the dependent variable was calculated as proportions of valenced and neutral material correctly recalled from the story, calculated by referring to a code scheme. As with Experiment 1, a series of mixed ANCOVAs was run for each dependent variable. Age, gender and trait anxiety scores were added as covariates in separate steps, and retained in subsequent steps only where they significantly predicted outcome measures.

Procedure

The procedure was identical to that of Experiment 1, with the addition of the SCAS completed immediately after the MASC-10, and the omission of individual clip introductions immediately prior to each clip.

Results

Familiarity with *Merlin*

The majority of children (*n* = 39 [69.95%]) had viewed *Merlin* on TV or DVD prior to the experiment, but were naïve to specific episodes from which the images were taken; only 15 children (26.79%) reported that they could recall seeing any part of the episode the monster images were taken from, and 9 children (16.07%) recalled any part of the episode the druid images were from.

Trait Anxiety

SCAS anxiety scores matched those from the normal population for 8-11 year olds (Spence, 1998), with normed *t*-scores ranging from 37-66 (*M* = 51.47, *SE* = 1.01, 95% BCa CI [49.25, 53.41]). There was no significant difference between male (*M* = 51.67, *SE* = 1.32) and female anxiety scores (*M* = 51.43, *SE* = 1.62), *t*(54) = 0.11 *p* =

.904, 95% BCa CI [−3.78, 4.34], *d* = 0.03, and age was not significantly correlated with anxiety, *r* = −.23, *p* = .094, 95% BCa CI [−0.47, 0.03].

Emotion Ratings

Children rated scores of happiness, sadness and fear for the images and narratives separately. A 2 (valence) × 2 (modality) × 2 (congruency) mixed ANOVA revealed that happiness ratings were significantly higher in the druid (*M* = 2.31, *SE* = 0.13), than the monster (*M* = 1.73, *SE* = 0.14) conditions overall, *F*(1, 53) = 11.70, *p* = .001, η_p^2 = .18, and for the narrative (*M* = 2.18, *SE* = 0.11) than the image stimuli (*M* = 1.87, *SE* = 0.13), *F*(1, 53) = 8.80, *p* = .005, η_p^2 = .14. A 2 (valence) × 2 (modality) × 2 (congruency) mixed ANOVA revealed that internalization ratings (sadness and fear combined) were significantly higher in the monster (*M* = 0.80, *SE* = 0.08), than the druid (*M* = 0.65, *SE* = 0.09) conditions overall, *F*(1, 52) = 6.29, *p* = .015, η_p^2 = .11.[9]

Recognition Task

Data from two children were removed from recognition analysis because their non-response rate was more than 2 *SD*s above the mean, the same approach as used in Experiment 1. There was no significant difference in response rates between the druid (*M* = 0.86, *SE* = 0.01) and monster conditions (*M* = 0.98, *SE* = 0.01), *F*(1, 52) = 0.42, *p* = .522, η_p^2 = .01, or between the congruent (*M* = 0.87, *SE* = 0.02), and incongruent conditions (*M* = 0.86, *SE* = 0.02), *F*(1, 52) = 0.07, *p* = .794, η_p^2 = .00. Table 1 shows the mean number of items recognized for each clip by memory type and modality.

[9] All other effects on happiness and internalization scores were non-significant and are reported in supplementary materials.

Table 1

Mean number of items recognized for each clip by memory type, modality and congruency of image and story valence, out of a maximum score of 6.

	Congruent Druid		Incongruent Druid		Congruent Monster		Incongruent Monster	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
True								
Visual	3.93	0.23	3.65	0.22	4.14	0.18	3.69	0.26
Narrative	4.64	0.19	4.69	0.19	5.39	0.14	5.15	0.21
False								
Visual	3.82	0.33	3.69	0.28	2.46	0.26	2.57	0.22
Narrative	3.68	0.38	3.80	0.28	4.39	0.27	4.46	0.24

Data from a third participant were removed because their *d*’ score in the negative visual condition was more than 3 *SD*s lower than the mean, the same approach used as in Experiment 1. A total of $n = 33$ boys and $n = 22$ girls were included in the recognition analysis, with $n = 27$ in the congruent conditions and $n = 26$ in the incongruent.

Sensitivity: *d*’ analysis. There was a high level of discrimination overall, with one sample *t*-tests showing that children could discriminate at better-than-chance levels in all conditions, with *d*’ scores significantly greater than 0. The following results were obtained in the congruent condition: druid visual $t(26) = 10.27$, $p < .001$, 95% BCa CI [1.16, 1.73], $d = 4.02$, druid narrative $t(26) = 8.56$, $p < .001$, 95% BCa CI [1.22, 1.92], $d = 3.36$, monster visual $t(26) = 8.80$, $p < .001$, 95% BCa CI [0.66, 1.02], $d = 3.45$, and monster narrative $t(26) = 13.34$, $p < .001$, 95% BCa CI [1.67, 2.30], $d = 5.23$.

In the incongruent condition, the following results were obtained: druid visual $t(25) = 7.13$, $p < .001$, 95% BCa CI [0.82, 1.37], $d = 2.85$, druid narrative $t(25) = 13.14$, $p < .001$, 95% BCa CI [1.72, 2.31], $d = 5.26$, monster visual $t(25) = 6.92$, $p < .001$, 95% BCa CI [0.62, 1.09], $d = 2.77$, and monster narrative $t(25) = 10.76$, $p < .001$, 95% BCa CI [1.31, 1.90], $d = 4.30$.

Age was significantly related to *d*' scores, $F(1, 50) = 6.84$, $p = .012$, $\eta_p^2 = .12$ and interacted significantly with story type, $F(1, 50) = 6.69$, $p = .013$, $\eta_p^2 = .12$. In the druid condition, discrimination increased with age for visual ($r = .26$, $p = .062$), and narrative stimuli ($r = .41$, $p = .002$), though the increase was only significant for narrative items. There was no significant effect of age on discrimination in the monster condition for visual ($r = -.03$, $p = .848$) or narrative stimuli ($r = .12$, $p = .411$). Gender and trait anxiety were not significantly related to *d*' scores (*p*s = .403 and .743). [10]

The resulting mixed 2 × 2 × 2 (Story [druid, monster], Modality [visual, narrative], Congruency [congruent, incongruent]) ANCOVA,[11] with age entered as a covariate, showed that children were significantly better at discriminating targets in the druid ($M = 1.53$, $SE = 0.08$) than the monster condition ($M = 1.33$, $SE = 0.07$), $F(1, 50) = 5.16$, $p = .027$, $\eta_p^2 = .09$. The difference between discrimination of narrative ($M = 1.80$, $SE = 0.08$), compared to visual targets ($M = 1.06$, $SE = 0.07$) was not significant, $F(1, 50) = 0.15$, $p = .702$, $\eta_p^2 = .00$, and neither was the interaction effect between story and modality, $F(1, 50) = 0.33$, $p = .857$, $\eta_p^2 = .00$.

There was no significant main effect of congruency, $F(1, 50) = 0.76$, $p = .388$, $\eta_p^2 = .012$, and no significant two way interactions between congruency and story, $F(1, 50) = 1.11$, $p = .297$, $\eta_p^2 = .02$, or congruency and modality, $F(1, 50) = 0.99$, $p = .324$, $\eta_p^2 = .02$. However, there was a significant story × modality × congruency interaction effect on *d*' scores, after accounting for age, $F(1, 51) = 9.46$, $p = .003$, $\eta_p^2 = .16$.

Figure 3 shows that in the congruent conditions, *d*' scores were highest for the monster (negative) narrative, ($M = 2.02$, $SE = 0.15$). The druid (positive) narrative ($M = 1.61$, $SE = 0.16$) and druid visual *d*' scores ($M = 1.48$, $SE = 0.14$) were similar to each

---

[10] See supplementary materials for full results of covariate analysis.

[11] Rotation was not entered as a fixed effect due to high levels of inter-correlation with congruency.

other, and the *d*' scores for the monster visual items (M = 0.84, SE = 0.11) were considerably lower. Paired *t*-tests revealed that the druid visual and narrative *d*' scores were not significantly different from each other, $t(26) = -0.56$, $p = .583$, 95% BCa CI [−0.48, 0.22], $d = 0.12$. However, in the monster condition, visual *d*' scores were significantly lower than narrative scores, $t(26) = -8.08$, $p < .001$, 95% BCa CI [−1.45, −0.89], $d = 1.82$. Children exhibited significantly higher levels of discrimination for druid compared to monster visual items, $t(26) = 3.46$, $p = .002$, 95% BCa CI [0.22, 1.00], $d = 0.77$, but no significant difference in discriminability was shown for druid compared to monster narrative items, $t(26) = -1.97$, $p = .059$, 95% BCa CI [−0.87, 0.02], $d = 0.39$.

In the incongruent condition, *d*' scores were highest in the druid (negative) narrative (M = 1.98, SE = 0.16), followed by the monster (positive) narrative (M = 1.59, SE = 0.15), then druid visual (M = 1.07, SE = 0.14) and monster visual condition (M = 0.86, SE = 0.11). Paired *t*-tests revealed that druid (negative) narrative *d*' scores were significantly higher than druid (positive) visual *d*' scores, $t(25) = -5.58$, $p = .001$, 95% BCa CI [−1.23, −0.60], $d = 1.08$, and monster (positive) narrative *d*' scores were significantly greater than monster (negative) visual *d*' score, $t(25) = 3.88$, $p = .001$, 95% BCa CI [0.35, 1.16], $d = 0.76$. There was no significant difference in discrimination between the druid and monster visual conditions, $t(25) = 1.33$, $p = .195$, 95% BCa CI [−0.12, 0.57], $d = 0.26$, but children were significantly better at discriminating druid (negative) compared to monster (positive) narrative items, $t(25) = 2.31$, $p = .030$, 95% BCa CI [0.09, 0.77], $d = 0.45$.

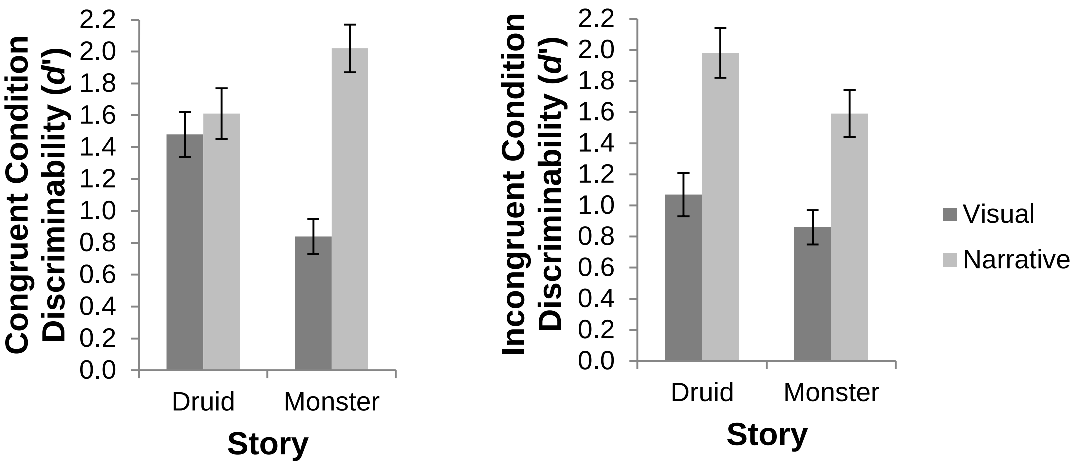


Figure 3. Interaction effects between type of clip and type of modality on discriminability for congruent and incongruent conditions (*d*' scores). Error bars show standard error.

Response bias: Criterion *C* analysis. In the congruent condition, *C* scores were significantly lower than 0 for monster visual, $t(26) = -3.53$, $p = .002$, 95% BCa CI [−0.48, −0.14], $d = 1.38$, and monster narrative items, $t(26) = -2.97$, $p = .006$, 95% BCa CI [−0.35, −0.08], $d = 1.16$, but not significantly different from 0 for the druid visual, $t(26) = 0.66$, $p = .516$, 95% BCa CI [−0.15, 0.27], $d = 0.26$, and druid narrative items, $t(26) = -1.74$, $p = .093$, 95% BCa CI [−0.41, 0 .02], $d = 0.68$. In the incongruent condition, they were also significantly lower than 0 for incongruent monster visual $t(25) = -3.07$, $p = .005$, 95% BCa CI [−0.43, −0.08], $d = 1.23$, monster narrative, $t(25) = -1.96$, $p = .061$, 95% BCa CI [−0.33, −0.00], $d = 0.78$, and druid narrative items, $t(25) = -1.92$, $p = .067$, 95% BCa CI [−0.28, −0.01], $d = 0.77$, when inspecting confidence intervals in addition to *p* values. *C* scores for druid visual items were not significantly different from 0, $t(25) = 0.21$, $p = .838$, 95% BCa CI [−0.17, 0.19], $d = 0.08$. This indicates liberal response bias for negative narratives, and negative images in both the congruent and incongruent condition, in addition to the positive narrative in the incongruent condition.

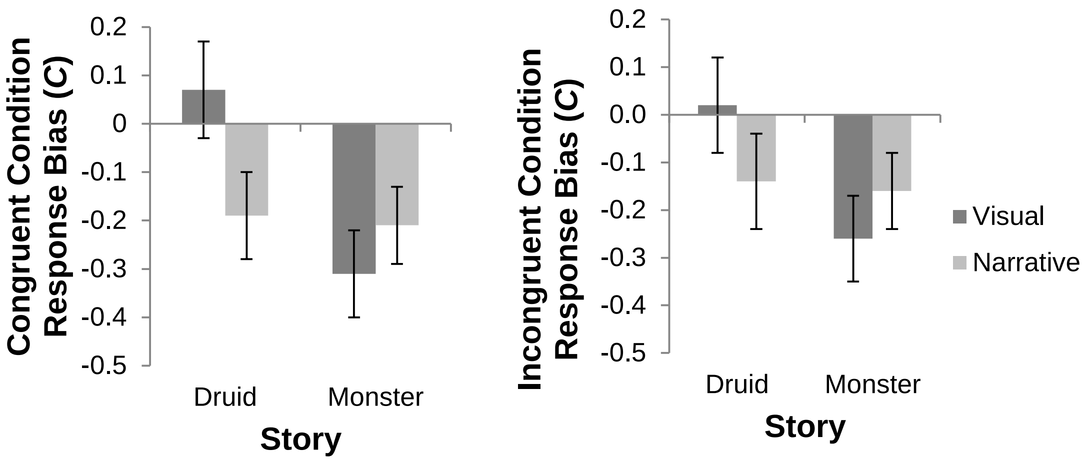


Figure 4: Interaction effects between clip modality on response bias, for congruent and incongruent conditions (*C* scores). Error bars show standard error.

Neither age, gender, nor trait anxiety were included as covariates in the main analysis because they did not significantly predict *C* scores (*p*s = .138, .180, and .566, respectively).[12] Criterion *C* scores for visual and narrative items across each story and congruency condition are shown in Figure 4. A 2 × 2 × 2 (Story [druid, monster], Modality [narrative, visual], Congruency [congruent, incongruent]) mixed ANOVA showed significantly greater liberal bias (bias towards answering 'yes') in the monster ($M = -0.23$, $SE = 0.05$) compared to the druid condition ($M = -0.06$, $SE = 0.06$), $F(1, 51) = 16.21$, $p < .001$, $\eta_p^2 = .24$. There was no significant difference in bias between narrative ($M = -0.17$, $SE = 0.05$) or visual items ($M = -0.12$, $SE = 0.05$), $F(1, 51) = 1.27$, $p = .266$, $\eta_p^2 = .02$, but the interaction effect between story × modality on bias scores was significant, $F(1, 51) = 8.92$, $p = .004$, $\eta_p^2 = 0.15$. Children exhibited significantly greater liberal response bias for visual monster ($M = -0.28$, $SE = 0.06$), than for visual druid items ($M = 0.04$, $SE = 0.07$), $t(52) = 2.24$, $p = .029$, 95% BCa CI [0.27, 0.53], $d = 0.62$, whereas levels of bias for narrative druid ($M = -0.16$, $SE = 0.07$)

[12] See supplementary materials for results of full covariate analysis.

and narrative monster items (*M* = −0.18, *SE* = 0.05) did not differ significantly, *t*(52) = −1.06, *p* = .295, 95% BCa CI [−0.38, 0.13], *d* = 0.04.

There was no significant difference in response bias between the congruent (*M* = −0.16, *SE* = 0.07) and incongruent condition (*M* = −0.13, *SE* = 0.07), *F*(1, 51) = 0.08, *p* = .776, η_p^2 = .00, and congruency did not significantly interact with story, *F*(1, 51) = 0.37, *p* = .548, η_p^2 = .01, modality, *F*(1, 51) = 0.27, *p* = .603, η_p^2 = .01, or story × modality, *F*(1, 51) = 0.26, *p* = .612, η_p^2 = .01.

Recall Task

Recall was transcribed verbatim from audio recordings of participants' responses. A code scheme was devised whereby the four narratives were chunked into distinct propositional items, in line with the approach used in Experiment 1. The three main visual features from each image were also included as items on the code scheme. This served as a checklist to calculate the proportion of emotional and neutral narrative, and visual material recalled from each story. Utterances were chunked and categorized using the following codes:

(1) Type of utterance: Narrative, visual or other. Each narrative item listed in the code scheme corresponded to a distinct proposition presented in the narrative, relating to an action, outcome, or unit of speech. The three main visual features for each picture were also listed as separate visual items within the code scheme. Other utterances included repetition, or anything unrelated to the narrative of the story or images viewed.

(2) Accuracy of narrative and visual utterances: True or false. This was determined by researchers consulting the narrative scripts and visual features included in the code scheme. Items were considered correct if they provided enough information to be matched against an item in the code scheme, without

containing false information. If they could not be matched to an item, or could be matched but contained faulty information, they were considered false.

(3) Valence of true narrative utterances: Emotional or Neutral. The two versions of stories for each set of images shared some identical narrative content (designated as neutral), and some unique valenced material. The unique, valenced text was highlighted on each script. The druid stories contained 71 neutral (shared) narrative items, with 55 valenced items in the positive, and 58 in the negative version. The monster stories contained 48 neutral items, with 71 valenced items in the positive, and 71 in the negative version.

Reliability of the code scheme was tested by an independent observer coding data for six children in the congruent and six children in the incongruent condition (21.43% of the sample). There was 88.94% inter-observer agreement (P_0) for utterance type (κ = 0.69), 93.35% for accuracy (κ = 0.50), and 89.49 % for valence of true, narrative utterances (κ = 0.79). Substantial Kappas are shown for both utterance type and valence (Landis and Koch, 1977), along with high P_0 values for all three categories. Although the Kappa coefficient for accuracy is only moderate, it is likely that the substantial imbalance in frequency of observed codes resulted in a disproportionately lowered κ (Feinstein & Cicchetti, 1990). The code scheme is therefore deemed reliable.

False memory utterances. Data from one participant were excluded due to the number of utterances exceeding three *SDs* of the mean in the druid condition. Age, gender and trait anxiety did not predict number of false memory utterances, so were not retained as covariates in the main analysis (*p*s = .446, .941 and .736, respectively).[13] The resulting 2 × 2 (Story [druid, monster], Congruency [congruent, incongruent]) mixed ANOVA revealed that frequency of false memory recall was not significantly

[13] See supplementary materials for full results of covariate analysis.

different between the druid (M = 1.06, SE = 0.16) and monster story (M = 1.12, SE = 0.16), $F(1, 51) = 0.27$, $p = .604$, $\eta_p^2 = .01$, or between the congruent (M = 0.87, SE = 0.21) and incongruent condition (M = 1.31, SE = 0.21), $F(1, 51) = 2.16$, $p = .148$, $\eta_p^2 =$ .04. The story × congruency interaction was not significant, $F(1, 51) = 0.27$, $p = .604$, $\eta_p^2 = .01$.

Visual utterances. There were also very few visual utterances across all conditions. As with false memory utterances, age, gender and trait anxiety did not predict the proportion of visual items correctly recalled, so were not retained in the main analysis (*p*s = .690, .819 and .290, respectively).[14] A 2 × 2 (Story [druid, monster], Congruency [congruent, incongruent]) mixed ANOVA revealed that the proportions of visual material correctly recalled for the druid (M = 0.19, SE = 0.05) and the monster condition (M = 0.24, SE = 0.07) were not significantly different, $F(1, 52) = 0.81$, p = .372, $\eta_p^2 =$ .02. There was no significant difference between the proportion of visual material correctly recalled in the congruent (M = 0.20, SE = 0.08) and incongruent conditions (M = 0.22, SE = 0.08), $F(1, 52) = 0.03$, $p = .865$, $\eta_p^2 =$ .00. The story × congruency interaction was also non-significant, $F(1, 52) = 0.81$, $p = .372$, $\eta_p^2 =$ .02.

True narrative utterances. Data from two participants were excluded; proportion of material recalled for one participant exceeded 3 *SD*s of the mean for neutral, and for valenced material, across both stories. Proportion of recall for neutral material in the druid story exceeded 3 *SD*s of the mean for the other. The proportion of true narrative material recalled from the valenced and neutral portions of each story is shown by congruency condition in Figure 5.

[14] See supplementary materials for full covariate analysis.

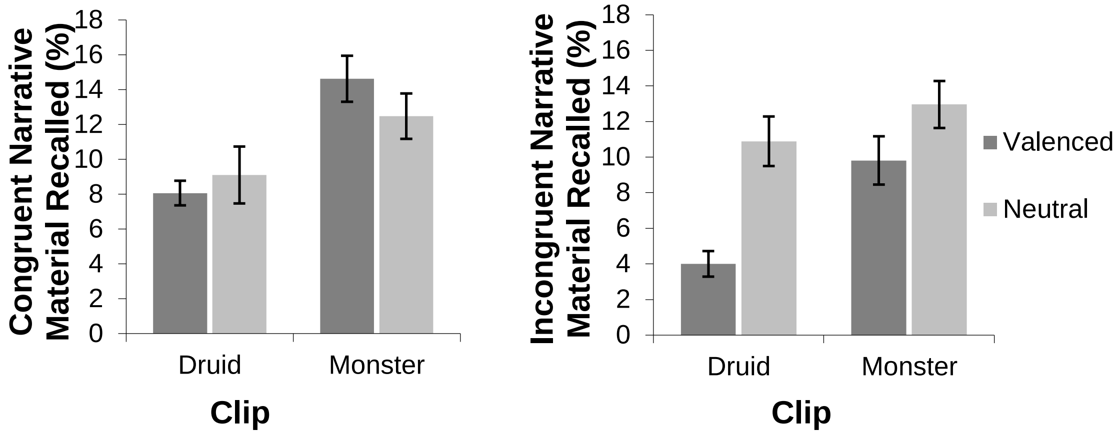


Figure 5: Interaction effects between clip and type of memory on percentage of true narrative items recalled in the congruent and incongruent conditions. Error bars show standard error.

Age, gender, and trait anxiety scores were not significantly related to the proportion of material recalled (*p*s = .097, .187, and .433, respectively).[15] A 2 × 2 × 2 (Clip [druid, monster], Material type [emotional, neutral], Congruency [congruent, incongruent]) mixed ANOVA was conducted. Overall, a significantly higher proportion of true narrative material was recalled from the monster (M = 0.13, SE = 0.01) than druid story (M = 0.10, SE = 0.01), $F(1,52) = 16.86$, $p < .001$, $\eta_p^2 = .25$. A significantly higher proportion was recalled from the neutral, shared portions of the stories (M = 0.13, SE = 0.01) than from the emotional portions (M = 0.11, SE = 0.01), $F(1,52) = 7.86$, $p = .007$, $\eta_p^2 = .13$. However, the interaction effect between story and material type was not significant, $F(1,52) = 0.11$, $p = .738$, $\eta_p^2 = .00$.

The main effect of congruency on recall was not significant, $F(1, 52) = 0.69$, $p = .410$, $\eta_p^2 = .01$, but did significantly interact with story, $F(1, 52) = 16.83$, $p < .001$, $\eta_p^2 = .24$, and material type on proportion of material recalled, $F(1, 52) = 14.62$, $p < .001$, $\eta_p^2 = .22$. Follow up analyses revealed that when the story and pictures were congruent in valence, recall was significantly greater in the monster (negative) (M = 0.12, SE = 0.01)

15 See supplementary materials for full covariate analysis.

than druid (positive) story condition (*M* = 0.09, *SE* = 0.01), *t*(27) = 2.24, *p* = .033, 95% BCa CI [0.01, 0.05]. When the story and pictures were incongruent, recall was significantly greater in the negative druid (*M* = 0.11, *SE* = 0.01), than the positive monster condition (*M* = 0.10, *SE* = 0.01), *t*(26) = −2.76, *p* = .010, 95% BCa CI [−0.01, −0.00]. When looking at which portions of the story were best recalled across both stories, there was no significant difference between the proportions of emotional (*M* = 0.12, *SE* = 0.01), and neutral material recalled (*M* = 0.11, *SE* = 0.01) when the images and story were congruent in valence, *t*(27) = 1.21, *p* = .235, 95% BCa CI [−0.01, 0.03]. In the incongruent condition, a significantly larger proportion of neutral material was recalled (*M* = 0.12, *SE* = 0.01), than emotional material (*M* = 0.11, *SE* = 0.01), *t*(26) = −5.79, *p* < .001, 95% BCa CI [−0.01, −0.01]. The three way interaction between story, material type and congruency on proportion of material recalled was not significant, *F*(1, 52) = 0.14, *p* = .712, η_p^2 = .00.

Discussion

Because valence was not successfully manipulated in Experiment 1 the clips are referred to as the ‘monster’ and ‘druid’, rather than negative and positive clips throughout the discussion. This aligns with the ‘story’ variable in Experiment 2, also referred to as the monster and druid story condition. The effects of Clip (Exp 1) and Story (Exp 2) on mood are discussed before moving onto the effects of valence and modality in memory, along with their interaction, and potential moderators of emotional memory. Implications and directions for future research are then considered.

Effects of Stimuli on Mood

To address the first two research questions, discrimination and recall performance for negative compared to positive materials is discussed. Valence was not successfully manipulated in Experiment 1, but the manipulation of valence was partially successful in Experiment 2. Overall, the druid condition evoked significantly greater happiness and lower internalizing emotions than the monster condition in Experiment 2. A significant three-way interaction would have been expected, should the manipulation have been fully successful. In each congruent condition, the paired image set and narrative should have been of equal valence, whilst the druid condition should have scored significantly higher on happiness and lower on internalization than the monster condition. In the incongruent condition, the image sets were paired with stories designed to be mismatched in valence, so the narrative in the druid condition should have been rated significantly lower on happiness and higher on internalization than the narrative in the monster condition (the opposite pattern to the congruent conditions). However, it was expected that the images would be rated the same between the congruent and incongruent conditions. None of these interactions were observed, which means that any memory effects resulting from an interaction between story condition, and modality or congruency are difficult to interpret.

It is possible that emotional rating scores were inaccurate because children rated the stories and images after completing the recall task for each condition. This resulted in a lag between presentation and self-report. Children were presented with the second set of materials before completing the recall and emotion rating tasks for the first set of presented materials, followed by the recall and emotion rating tasks for the second set of materials presented. Likewise, the fact that the two stimuli sets (narrative and image sets) were presented simultaneously is likely to have impacted overall emotion evoked in the condition, possibly confounding results for the individual components. Power is also likely to have been an issue, with fewer than 30 children each in the congruent and incongruent condition. The individual narratives and image sets should have been presented singularly to a different sample of children, for valence ratings to be obtained prior to use in the main experiment. Because these ratings may not be fully reliable, modality and congruency effects have been interpreted as originally intended, with the incongruent druid narrative labelled as negative, and the incongruent monster narrative labelled as positive.

Additionally, it should also be noted that happiness ratings were significantly lower for images than narratives for both story conditions, although the effect of this was much larger for the druid condition. Ideally, strength of valence should be equal for images and narratives to compare the effect of valence between modalities.

Valence Effects on Memory

Findings relating to research aims (1) identify any differences between true and false memories for negative and positive materials in children; and (2) explore whether any differences in true and false memories for valenced material vary, depending on type of stimulus, are discussed in this section, along with the interaction between valence effects and context on memory. Overall, no significant effects on recognition or correct recall were observed in Experiment 1. The significant valence effects in Experiment 2 are as follows: (1) a significant interaction effect of valence and modality on memory, with more accurate discrimination and higher correct recall for negative than positive narrative material; (2) a significant interaction effect of valence, modality, and congruency on discrimination, with higher accuracy for positive than negative visual stimuli in the congruent condition only; and (3) lower rates of recall for emotional than neutral narrative material, significantly so when the story and images were incongruent.

Finding (1) carried across recognition and recall tasks, as well as the congruent and incongruent conditions. No differences in response bias for negative and narrative content were observed, either in the congruent or incongruent condition. This finding further highlights that there is a genuine difference in children's ability to discriminate both positive and negative stimuli from noise effectively. The valence effect on discrimination was larger in the congruent than incongruent condition. In the incongruent condition, it is possible that the difference between the valence of the image set and its matched narrative is decreased, with valence ratings becoming relatively more similar to each other, because the images appear to show a similar outcome to the story. This is known as acquired equivalence, where generalization is increased between two superficially dissimilar stimuli (or antecedents) that have previously been associated with similar outcomes (or consequents) (Meeter, Shohamy, & Myers, 2009). Where this has happened, the valence of the narratives in the congruent conditions will have become weaker, resulting in a potentially weaker effect of a difference between them.

Greater discrimination for positive, than negative visual stimuli was only identified in the congruent condition. The same effect was not found for recall of visual items. Nevertheless, it was a medium effect size and can be interpreted in the context of response bias. Children showed lower levels of liberal response bias towards the positive than negative visual stimuli. This means that as well as being able to better detect signals from noise, they were less likely to be biased by another decision criterion. Despite a lack of valence or modality effects on discrimination in Experiment 1, the same pattern of response bias was found for visual items, though only when the negative clip was viewed first.

Findings (1) and (2) are broadly consistent with previous, but sporadic research findings of higher discrimination for positive than negative material of a less complex nature (Brainerd et al, 2010; Moradi et al., 2000), and greater recall for negative compared to positive material of a complex nature (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015).

However, congruency also appears to interact with modality (independently of valence). When incongruent images and stories were presented, discrimination was greater for the narratives than images, regardless of what valence the narratives were. Presenting contradictory valenced material simultaneously could have triggered cognitive dissonance, resulting in a state of mental stress or discomfort (Festinger, 1957). Threat processing literature shows three main categories of coping behaviour during a stressful situation; blunting (including reappraisal, cognitive distraction and behavioural distraction), monitoring (including mentally focusing on aspects of the threat situation, attending to threat cues and seeking information) and seeking support/ comfort (Hoffner, 1993; Miller, 1987). When presented with stressful stories, children of a similar age most commonly report mentally focusing on aspects of the threat situation (Hoffner, 1993). Despite this being the most reported individual behaviour; blunting behaviours were reported more frequently when combined into one overarching category. It is possible that children were employing some of these strategies in Experiment 2. Children may have been less inclined to perform cognitive distraction because they were told they would be discussing the stories afterwards, and therefore needed to pay attention to the presented stories and images. There was also limited opportunity for behavioural distraction because there were no other children or materials to engage with nearby. There was little opportunity to seek comfort or support as only the interviewer and child were present, and the children were wearing headphones. This means that a monitoring coping style may have been adopted more frequently, with children focusing more on particular details of the story, seeking information and attending to the threat cues in the narratives over the visuals. As outlined by Kensinger, (2004) an increase in attention to valenced stimuli can increase memory performance. Measures of attention would need to be included should this be experiment be replicated or extended upon in the future, to explore this hypothesis.

It is surprising that the effect of stimulus modality on discrimination or recall had not also been found in Experiment 1, because the same narrative and visual questions were used with the clips containing visual (moving) images, and narrative material (including speech from the characters). However, visual and narrative items may not always have been completely distinct. For example, where you could both see a character perform an action and the characters were talking about the action.

Further, negativity and positivity preferences could be tested in Experiment 2 with the addition of neutral portions of narrative. The recall task showed a preference away from emotional narrative material, but only in the incongruent condition. Although several existing papers also report a preference away from negative and positive words (Howe, 2007; Howe et al., 2010; Moradi et al., 2000), these all used word lists. This finding is therefore surprising, given Van Bergen et al (2015) found a preference toward negative and positive, compared to neutral material when stories were used.

Evolutionary theories that postulate only threatening material is privileged (for example, Öhman & Mineka, 2001; Tooby & Cosmides, 1990) cannot be tested in Experiment 1, due to the lack of a neutral condition. There was also limited scope for testing in Experiment 2 because the neutral material was integrated into both positive and negative narratives, rather than including a discrete neutral condition. The same discrimination questions were used between Experiment 1 and 2, which means discrimination was not tested for neutral portions of the story.

The general pattern of valence effects does not lend support to evolutionary models which propose that all evolutionarily advantageous material is privileged in cognitive processing (for example, Capiaoppo et al, 1997;1999; Hamann, 2001), or cognitive models proposing all valenced material is processed preferentially, for example the Network Affect Model (Bower & Cohen, 1982), or cognitive elaboration models (Kensinger, 2004).

False Recall

The number of false memories recalled in Experiment 1 also differed significantly between clips, although this depended on which clip was presented first; no such differences were observed in Experiment 2. In both presentation orders, the second clip elicited greater false recall than the first (regardless of which valence clip was presented second). This medium effect was only significant when the druid clip was presented first, with more false memories of the monster clip recalled.

This initially suggests a genuine interaction effect, reflecting an aspect of memory distinct from emotional memory, such as a recency effect. Existing literature has identified serial position effects in false memory recall but found that a recency effect was observed only in specific circumstances; for phonologically, but not semantically related lists of words, and that this effect was diminished when a delay was added into the paradigm (Lane, 2010; Zucker, 2010). Within this task, material was semantically related and a delayed recall task was employed, which does not agree with previous findings. However, the stimuli type is not comparable and it is possible that another aspect of the clips may be moderating the serial position effects. Additionally, rates of false recall for the monster clip did not vary between presentation order, but there was a large effect of presentation order on rates of false recall for the druid clip. It therefore appears only false memories for the druid clip were susceptible to presentation order. Any unintentional difference between the clips would have impacted these results. This would warrant further investigation to ascertain which unintentional differences are evident between the two clips, and whether these had an effect on false recall rates and the apparent recency effect. Despite the large effects, the rate of false recall was comparatively low when examining true recall rates. Without further analysis it is unclear whether rates of false recall would affect recall accuracy rates,[16] as determined by dividing the frequency of correct items recalled by the total number of items recalled (Jou, 2008).

Moderators of Emotional Memory

Moving onto research aim (3) age, gender, and trait anxiety were explored in regards to their overall relationship with memory performance, and to identify whether this differed between valence and stimulus type. In general, there was some evidence of memory performance increasing with age, whilst only one effect of gender, and one interaction effect of trait anxiety was observed across all analyses.

Age. Overall discrimination and true narrative recall for negative material increased with age in Experiment 1, with medium effect sizes observed. A smaller positive age effect on discrimination was observed in the druid condition of Experiment 2. A positive relationship is in line with existing research and theories asserting that long term memory ability increases with age from early childhood into adolescence, although the rate of development starts to slow down from around age 7 years (Gathercole, 1998). In addition, a small positive effect of age on number of false memories was observed in Experiment 1. These findings are all in line with results from other emotional memory experiments with children (Howe, 2007; Howe et al., 2010,

---

[16] False recall is measured in number of utterances for each clip overall, whereas true recall is measured as a proportion of material recalled from positive and negative scenes of each clips. False recall items would therefore need to be re-coded as belonging to negative or positive scenes, and weightings applied to account for different proportions of materials presented in positive and negative scenes in each clip.

2009; Howe & Wilkinson, 2011). An increase in recognition and recall was expected for all conditions and tasks, across both experiments. It is unclear why this effect wasn't found reliably across Experiment 2, particularly as the sample size was larger than Experiment 1. There were at least 26 participants per cell, above the minimum suggested by Simmons et al. (2011) but still marginally below the figure suggested by Cohen (1988).

Gender. Gender co-varied with levels of recognition in Experiment 1 with males discriminating correct answers more accurately than females. However, the reason for this gender effect is not immediately clear. *Merlin* is a fantasy TV show; the fantasy genre is typically thought of as appealing more to boys than girls (Streicher & Bonney, 1974). Although both stimuli sets were derived from the same TV show, Experiment 1 used clips edited directly from the episodes, which were more in keeping with the original programme format. The clips may therefore have a stronger fantasy element than the stimuli used in Experiment 2, explaining why this effect was not replicated. Gender effects were not found in response bias or recall, and there were no significant gender differences reported in frequency of viewing *Merlin* at home, how much boys and girls liked watching *Merlin*, or levels of excitement after viewing the clips in Experiment 1,[17] suggesting that gender effects were not due to a male preference for fantasy TV. Previous literature suggests females tend to have a slight memory advantage in general, the opposite to that which was found here (Herlitz et al., 1997). In addition, no effect of gender has been identified on recognition scores in child DRM studies (Cordon et al., 2013; Howe et al., 2010; Howe & Derbish, 2010). Gender did not interact with valence effects and so is not a central finding to the thesis. This effect has not been further explored, and is treated as a spurious finding.

[17] See supplementary materials for excitement ratings analysis for Experiments 1 and 2.

Trait anxiety. Trait anxiety interacted with valence and clip on the proportion of true material recalled in Experiment 1. Recall of material from the dominantly valenced portion of each clip increased with age, whilst there was no difference for the minority portions. This did not match the expected interaction, whereby recall would increase with trait anxiety for the negative clip, but decrease with trait anxiety for the positive clip in line with adult research (Mitte, 2008). This finding was not seen across other analyses, and due to both experiments being underpowered to a greater or lesser extent, it is difficult to ascertain whether there is a genuine null effect of trait anxiety on memory for emotionally negative stimuli, or stimuli in general. Future research should seek to achieve an appropriate level of power to determine whether reliable true effects of anxiety do in fact exist. Consequently, it was not possible to test predictions of developmental theories of fear acquisition (Field & Lester, 2010). Comparisons to previous empirical findings were also limited.

Conclusions

There were several key findings in relation to the original research aims. Aim (1) was to identify any valence effects on true and false memories in typically developing children. Memory performance was generally better for negative than for positive narrative material. This effect may signify a plausible mechanism behind the internalizing effects of scary TV on children’s emotions, although future studies should seek to include separate, piloted, positively and negatively valenced, and neutral TV clips, to provide more robust evidence of this.

Research aim (2) was to explore whether valence effects in memory differed depending on stimulus modality. Discrimination was greater for narrative than visual items in all conditions, with an interaction observed between modality and valence. Effects of modality were stronger when at least a portion of the material presented was negative. One plausible explanation for this is that children employed a monitoring coping style response, with children seeking information from the richest source available when faced with a stressful event or situation (Hoffner, 1993; Miller, 1987). This was paired with a preference away from valenced material, and towards recalling more neutral material in the incongruent conditions. This could suggest that children seek to process the neutral portions of the story to a greater extent to gain further information, in order to ease the state of cognitive discomfort caused by cognitive dissonance. When considering TV viewing, this suggests that children are likely to remember story lines more than images. Parents should ensure that this particular aspect of viewing is closely monitored to prevent their children from viewing material that they might find distressing.

Although the interaction between modality and story was part of a three-way interaction with congruency, it is difficult to ascertain what the congruency effect really means because emotional ratings did not differ between these conditions as anticipated.

None of the moderators explored appeared to interact with valence (research aim 3), despite a tendency for general memory performance for true and false recall, and recognition to increase with age. Trait anxiety and gender effects were limited, and considered potentially unreliable.

Future experiments should ensure that all stimuli sets are piloted thoroughly with the target audience and adapted accordingly, should seek to explore what feature, or features of the clips were interacting with modality and presentation order to produce these effects and ensure adequate power by running power calculations in advance where possible. The interaction effects with congruency would be a theoretically interesting next step for research.

Chapter IV

Effects of Emotional Context and Valence on Children's True and False Memories

Abstract

The role of emotional context and valence on children's true and false memories was examined using the Deese-Roediger-McDermott paradigm. Eighty-four 7-11 year old children had an emotionally positive, neutral or negative story presented to them as a mood inducer. Lists of emotionally positive, neutral and negative words were then presented, each followed by an immediate recognition task. Contrary to previous findings, discrimination was greater for emotionally negative than positive words; only when they were presented first. A corresponding decrease in conservative response bias was observed for emotionally negative words in these conditions. Discrimination was expected to be greater for the word lists which matched the valence of the story, but this effect was not borne out. Discrimination findings are likely due to differences in arousal and age of acquisition between word lists, and methodological differences between the current and previous experiments. Consequently, limited conclusions can be drawn.

Effect of Emotional Context and Valence on Children's True and False Memories

For many years, theories regarding the processing of emotional material have pointed towards a phenomenon known as the positive-negative asymmetry (Lewicka et al., 1992). Positivity offset refers to the tendency to evaluate situations or stimuli positively, particularly if they are presented in a neutral light; in addition to a preference towards mildly positive, compared to mildly negative stimulus within cognitive processing such as memory and attention (Czapiński, 1982). Negativity 'bias' (preference) results because the activation function for negative stimuli has a higher gain than the activation function for positive stimuli, meaning this preference is shown for more strongly emotional stimuli. Negativity preference can be broken down into affective and informational sub-types. Affective negativity preference means that negative stimuli will have a larger impact on affect and behaviour than positive stimuli of the same intensity, whereas cognitive negativity preference means a higher level of cognitive elaboration is seen for negative than positive stimuli of the same intensity (Lewicka et al., 1992).

Negativity preferences in recognition (younger adults) and recall (younger and older adults), and positivity preferences in recall (younger and older adults) are evident across the general population, based on the results of a meta-analysis (Murphy & Isaacowitz, 2008). The negativity effects are larger than the positivity effects. The opposite effect is generally found in children when a DRM paradigm is used with word lists; preferences away from negative and positive words are observed in both recall and recognition tasks (Howe, 2007; Howe et al., 2010; Moradi et al., 2000; Neshat-Doost et al., 1998). There is only one notable exception to this; when images are used instead of word lists, negativity and positivity preferences are shown in recall memory (Cordon, 2013; Van Bergen et al., 2015). As with adults, negativity preferences are also larger than positivity preferences in recall (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015). This indicates the pattern for recall of narrative items may be generally similar between children and adults, but differ for recognition.

The majority of experiments exploring children's emotional memory are DRM studies using word lists, and do not often include positive stimuli. Experiments 1 and 2 of this thesis therefore sought to: (1) compare memory performance for negative and positive material directly; (2) attempt to replicate previous findings using narrative and visual stimuli; and (3) explore age, gender, and trait anxiety as potential moderators of valence effects on memory. TV clips were used in Experiment 1. Due to a desire to keep the stimuli as naturalistic as possible with a coherent storyline, each valenced clip contained a small portion of its counterpart. Valence was not successfully manipulated, and these aims could not be tested. Experiment 2 attempted to build on this design by using still images from the clips in Experiment1, and including both a positive and negative narrative for each. This aim of this was to reduce the confounding variables by including two instances of each valence story. In line with previous research, negative narrative material was discriminated at a greater rate than positive. However, when the valence of the images and narrative was incongruent, narrative material was also recognized more accurately than visual, regardless of the valence. Conversely, positive visual material was discriminated more successfully, significantly more so when valence of images and narrative was congruent. This is also in line with previous findings.

Context Effects on Emotional Memory

In addition to differences in memory performance based on modality and complexity of stimuli, the context surrounding encoding, and recognition or recall must also be considered as a moderating factor. This is especially important when considering naturalistic studies, where single effects cannot be studied in isolation. Bower's Network Theory of Affect specifies that memory performance is enhanced when mood at learning or encoding matches that at retrieval (mood dependent memory), when the stimulus valence matches mood during learning, and when stimulus valence matches mood during retrieval (mood congruent memory) (Singer & Salovey, 1988).

Many studies into mood-state and mood congruency have been conducted in both clinical and non-clinical adult samples. Meta-analyses including non-clinical adult samples have shown that memory performance is significantly enhanced when positive mood (Ucros, 1989) and negative mood (Matt, Vázquez, & Campbell, 1992; Ucros, 1989) is induced at both encoding and retrieval, although a review reports that this effect is not reliably found in children (Blaney, 1986). In its simplest form, this provides evidence of environmental context-dependent memory, where affect is part of the learning environment (Smith & Vela, 2001). However, it could also be interpreted as state-dependent learning, where memory performance is enhanced when physiological state is the same at encoding and retrieval (Petersen, 1979). For example, happiness, sadness, anger and fear were all shown to affect systolic and diastolic blood pressure, and heart rate differentially after exercise (Schwartz, Weinberger, & Singer, 1981).

An early meta-analysis which included mostly non-clinical adults found a significant effect of mood-state congruency on retrieval performance, when comparing congruent contexts where output and item valence match, to incongruent contexts, $z = 0.44$, CI[18] [0.26, 0.62] (Ucros, 1989). There was a significantly stronger effect in positive than negative mood-congruent conditions. When effects are broken down by study design, the mood congruency effect was biggest for lived-events, followed by intentional and then incidental learning. A subsequent meta-analysis including mostly

---

18 Size of confidence was not reported by Ucros (1989).

non-clinical adult samples also found a small but significant effect of better recall of positive than negative stimuli overall, weighted mean effect $\overline{d}_h$ = 0.15, 90% CI [0.08, 0.22] (Matt et al., 1992). When depressed mood was induced, *d* = −0.12, 90% CI [−0.19, −0.04], and when elated mood was induced, *d* = 0.08, 90% CI [0.01, 0.05]. The effect size was larger for depressed mood, and confidence intervals suggest the true population mean was likely not to be 0 for negative-mood congruency only. This analysis suggests a significant mood-congruency effect for negative but not positive mood congruency, contrary to findings by Ucros (1989).

A limited number of studies with child and adolescent samples have been published. Amongst them, Bartlett, Burleson and Santrock (1982) compared congruency of stimulus valence and mood at recall and found that children recalled more happy words when they were happy during recall, compared to when they were sad at recall. There was no significant difference between the frequencies of sad words children recalled depending on whether they were happy or sad during recall. Nasby and Yando (1982) also found a significant interaction between mood at recall and word valence. Children who were happy at recall retrieved more happy words than those who experienced a neutral mood, and those who were sad at recall retrieved fewer happy words.

More recently, the effects of input-mood and item-valence congruency on recall has been studied in children (Bishop, Dalgleish, & Yule, 2004). A non-clinical sample of children was split into low and high depressed groups based on the Depression Self Report Scale (DSRS; Birleson, 1981). Positive, neutral and negative versions of three stories were presented followed by a recall task. Children in the low depressed group recalled significantly more from the positive than neutral story, and more in the negative than neutral story, with the difference approaching significance. There was no significant difference between the amounts of information recalled from the negative compared to the positive stories though. However, children in the high depressed group recalled significantly more from the negative than the positive or neutral stories. From this study, it is suggested that a depressed mood at input leads to a preference for recall of congruently valenced, negative material in children. It is arguable whether children who were categorized as high on a depression scale would necessarily be in a negative emotional mood specifically during the experiment, however.

Most previous emotional memory studies have focused on memory preferences in anxious or depressed samples. Negativity preferences have been shown in clinically and non-clinically anxious and depressed children and adolescents in the majority of existing literature (Bishop et al., 2004; Daleiden, 1998; Dalgleish et al., 2003; Dalgleish et al., 2001; Ladouceur et al., 2005). However, results are mixed in control samples of children, with a negativity preference shown in one study (Baltazar et al., 2012), a positivity preference in other studies (Bishop et al., 2004; Ladouceur et al., 2005) and a preference away from emotionally negative stimuli in some existing DRM studies (Howe et al., 2010; Roediger & McDermott, 1995). Previous studies employed a variety of paradigms and contexts, which may explain some of these inconsistencies. Baltazar et al. (2012) used photographs of characters that had good or bad actions or attributes associated with them. During the recognition phase, children were asked closed questions about the characters or events. In contrast, Bishop et al. (2004) used a valenced story as stimuli but employed a recall task rather than a recognition task, and the Ladouceur et al. (2005) study employed an E-*n*-back task, which focused on working memory rather than short or long term memory. Finally, Howe et al. (2010) and Roediger and McDermott (1995) employed a DRM paradigm utilizing single word items without any social context or images attached to them. It may be that a negativity bias is shown more in a social context with recognition tasks, positivity bias during working memory and for recall tasks, but a bias away from emotional content when stimuli consists purely of singular words with no visuals or social context.

The current study aims to improve and extend on existing studies into children's memory for emotionally salient stimuli by exploring negativity and positivity preferences in an immediate recognition task, and to further explore context effects between mood (induced by a valenced story) and discrimination for valenced word lists in a non-clinical child sample. Children will complete immediate recognition tasks after viewing positive, neutral and negative emotional DRM lists. This allows for both true and false memory to be measured. Positive lists are included to tease apart the relationship between positivity and negativity preferences.

Prior to the recognition tasks, positive, neutral or negative affect will be primed using the nine versions of valenced stories created by Bishop et al. (2004) to explore the effect of the congruency of input mood and stimulus affect. Using valenced stories to provide an emotional context will enable exploration of the effects of context on emotional processing bias. In Bishop et al.'s original study, the stories were not used to manipulate mood; children were split into low and high depressed groups based on whether their self-report scores of depression were above or below the median score within the sample. This method of forming groups for comparison is not ideal because it relies on self-report which can be especially unreliable in younger children (Schwab-Stone, Fallon, Briggs, & Crowther, 1994). Younger children may also lack a frame of reference to compare mood or symptoms to normative characteristics (Tarullo, Richardson, Radke-Yarrow, & Martinez, 1995). The focus of the Bishop et al (2004) study was also slightly different; the authors were interested in examining emotional memory specifically in relation to depression, rather than mood congruency.

Consequently, there was no neutral or positive mood comparison group and no guarantee than those scoring high on depression were currently experiencing a negative state.

As in Experiments 1 and 2 of this thesis, age, gender and trait anxiety will be explored as potential moderators of emotional memory. Trait anxiety will also be measured using the SCAS (Spence, 1997; 1998). Bishop et al. (2004) found an effect of levels of depression on mood congruency and emotional processing bias, in a non-clinical sample of children. Inclusion of the SCAS will allow comparison of the effects of anxiety levels on mood congruency and emotional processing bias in a non-clinical sample. Word list order of presentation was rotated as a counterbalance measure but included in analysis as a fixed effect because Experiment 1 identified some potential differences in memory performance based on rotation order. In Experiment 1, there was significantly greater liberal bias for the negative visual items compared to positive visual items, but greater liberal bias for the positive narrative items than negative narrative items, only when the negative clip was presented first. Additionally, significantly more false memories of the monster (negative) than the druid (positive) clip were recalled, when the druid clip was presented first.

The aims of this experiment are: (1) to explore negativity and positivity preferences in an immediate recognition task; (2) to explore context effects between mood (induced by a valenced story) and discrimination for valenced word lists; and (3) to explore potential moderators of emotional memory in children (age, gender, trait anxiety).

Based on previous research, it is hypothesized that: (1) negative words will be discriminated at a lower rate than neutral words; (2) words matching the valence of the story will be discriminated at a higher rate than those that do not; (3) discrimination will increase with age; and (4) there will be no gender effect on discrimination.

Method

Participants

Three schools in East Sussex and Surrey, England participated in the study. Ninety-four opt-in consent forms were returned by parents of children in school years 3-6 (age 7-11 years). Inclusion criteria required all children to speak English as their first language because the memory tasks used English words as stimuli, and additional languages are processed differently to primary languages (Hahne, 2001). Parents were asked not to return consent forms if their child had a formal diagnosis of behavioural, developmental or educational disorders, to limit any undue task difficulty or stress felt by participants. All children were also required to have normal, or corrected to normal hearing and vision because stimuli was presented visually and aurally. Only one child per family was allowed to participate to limit shared sources of variance between participants. Seven children were excluded because their siblings had taken part, one did not provide assent, one requested to leave the interview mid-way and one did not meet inclusion criteria based on teacher report. A total of 84 children completed the experiment. Forty participants (47.62%) were boys and 44 (52.38%) were girls. Ages ranged from 7 years, 6 months to 11 years, 6 months (M = 9 years, 5 months, SD = 13.08 months). The group was ethnically, culturally and socio-economically mixed. Pupils were recruited from a mixed Community State School with a lower than average proportion of pupils eligible for the pupil premium, a larger proportion of White British pupils, and an Ofsted rating of ‘requires improvement’, a mixed voluntary aided Catholic School with a lower than average proportion of pupils eligible for the pupil premium, a higher than average proportion of pupils from a minority ethnic background and an Ofsted rating of ‘good’, and a mixed academy primary school. Children took part voluntarily and were naïve to the purpose of the experiment. Children were entered into a voucher prize draw on completion.

Materials

As with Experiment 2, The SCAS (Spence, 1997; 1998) was administered to measure levels of trait anxiety. The SCAS is a 45-item scale measuring six domains of anxiety, and includes six filler items. Muris et al. (2000) found evidence to support the psychometric properties of the SCAS in 7-11 year olds. Items were read aloud to children, who were asked to circle how often they experienced each one on a 4-point scale (never = 0, sometimes, often or always = 3). A total score was calculated by summing responses of the 39 anxiety items and converting them into a *t*-score using a chart provided by the author. The SCAS appeared to have good internal consistency in the current sample, $\alpha = 89$.

Nine variations of narratives, based on three stories were audio recorded for use in the study (Bishop et al., 2004). The three stories were about a child at the beach, a child at the park, and a child coming home from school. There were three versions of each story; a positive, neutral and negative version. A male and female version of each of the nine story variants was recorded with different character names and pronouns used, as well as a matching male or female narrator, to match the gender of the participant it was being presented to. The stories consisted of 7 cartoon images, each displayed for 23 s with the narrative recording played through headphones simultaneously. The stories were presented in MS PowerPoint using a laptop. The images remained the same for each valenced version of the same story, and consisted of the same narrative for the first three slides with the narrative for the last four slides varying between versions. The three versions of the beach story averaged 428 words in length, the three versions of the park story averaged 455 words, and the three version of the coming home story averaged 396 words. All visual and audio files were played at a standardized brightness and volume through the same laptop. A practice story consisting of three cartoon images was also prepared for presentation prior to the experimental story. The practice images were each displayed for 23 s, with a recording of a narrative describing some children picking apples.

There were six DRM stimuli lists, each consisting of 14 items (see Appendix: Chapter IV, V – Stimuli). Items were taken from 12 previously established backwards associative strength (BAS) DRM stimuli lists, with items 2-8 from each list included in the experimental stimuli lists. List A and B were made of positive items (A [‘Sweet’ and ‘Love’], B [‘Soft’ and ‘Music’]), C and D from neutral items (C [‘Green’ and ‘Chair’], D [‘Fruit’ and ‘Foot’]) and E and F of negative items (E [‘Cry’ and ‘Thief’], F [‘Anger’ and ‘Sick’]). The lists Sweet, Chair, Fruit, Foot, and Cry were sampled from lists used by Budson et al. (2006), The lists Love, Soft, Music, Green, Thief, Angry, and Sick were sampled from the Cornell/ Cortland emotion lists (Brainerd et al., 2010). There was no overlap between items across lists.

The order of items was randomized within each of the 12 previously established lists, with items from each list pair presented alternately to make the final six lists. The order of item presentation in each list remained fixed for all children. Each item was presented in bold black font, size 115 on a beige background, with a recording of the word read out via headphones. Each item was displayed for 2 s. Each recognition task list consisted of a presentation of 20 items. Two were the not-presented critical lures from the stimuli lists, two were the first items from each established DRM list which had not been presented (related lures), 10 were unrelated distractors from alternate DRM lists of the same valence and six were target items (items from position 3, 5 and 7 in each presented DRM stimuli list). The order of items was randomized and presented in a fixed order. Each item was displayed on screen in bold black font, size 115 with a recording of the word played over headphones. Children were asked to say ‘yes’ out loud if they remembered the word appearing on the previous list or ‘no’ if they did not. The researcher recorded their response and displayed the next item.

Children rated emotions before and after listening to the story on a 4-item self-report questionnaire using a 5-point Likert Scale to state how strongly hey felt each emotion, from ‘0’ (Not at all) to ‘4’ (Very, very much). The four emotional responses recorded were: feeling scared, happy, sad and excited. They were also asked whether they enjoyed the story, and found it interesting, using the same scale.

Design

First, manipulation checks for mood induction by the story were conducted. Story type (beach, park, coming home), and Story valence (positive, neutral, negative) were the independent variables. Both of these were between-participant variables. The dependent variables in the manipulation checks were ratings of happiness, and ratings of internalization. Because there were no significant differences in scores of happiness, and scores of internalization depending on story type, scores were collapsed across the three positive, three neutral, and three negative conditions. Story type was therefore not entered as a variable in the main recognition analyses.

A mixed design was used for the main recognition analyses. Each of the two independent variables had three levels: Valence of story (positive, neutral, negative), and Valence of word lists (positive, neutral, negative). Rotation order of word lists (1-6) was also entered as a fixed effect in analysis. Valence of story was a between-participant variable; valence of word list was within-participant. The two dependent variables for recognition analysis were hit and false alarm rates, used to calculate *d*’ and

C scores. Gender, age, and trait anxiety scores were included as covariates in separate steps, but retained only if they significantly affected recognition scores.

Procedure

Interviews were conducted individually in a quiet space at the child's school, away from their classroom. Children had the interview procedure and confidentiality information outlined to them. They were given a chance to ask questions and assent was attained. For all self-report measures the researcher read the questions out loud, and the children circled their response on the answer sheet. Children were able to ask questions if they did not understand any of the items.

Children completed the SCAS. They watched the practice story followed by one experimental story, and then rated emotions evoked by the story. Order of story presentation was rotated through the nine story versions, by story type (beach, park, coming home) and valence (positive, neutral, negative). For example, the first child was shown the positive-beach story, the next child was shown the neutral-beach, and the next the negative beach story, before moving through the three park, and then three coming home story versions. The male version of the respective experimental story was presented to boys, and the female version was presented to girls.

Each word list was presented in turn, immediately followed by the relevant recognition task. This was repeated for the six word lists. Pairs of word lists of the same valence were blocked, with the order within pairs counterbalanced; A-B, C-D, E-F. The order of presentation was also rotated by valence; positive, neutral, negative. Finally, children were asked how much they enjoyed the story and found it interesting, were debriefed, thanked, awarded a sticker, and given the option of viewing the correct recognition answers before returning to class.

Children were monitored closely for any signs of distress, such as covering eyes or engaging in distraction behaviours. No children showed any visible signs of distress, but if they had, the interview would have been terminated immediately with children being given the choice of reading an age appropriate book before returning to class.

Results

Trait Anxiety

SCAS anxiety scores matched the normal population for 8-11 year olds (Spence, 1998) with normed *t*-scores ranging from 30-68 (*M* = 50.00, *SE* = 1.00, 95% BCa CI [47.95, 51.81]). A 2-way ANOVA revealed no significant difference between male (*M* = 51.53, *SE* = 1.51) and female SCAS scores (*M* = 48.84, *SE* = 1.43), *F*(1,78) = 1.47, *p* = .229, η_p^2 = 0.02, or between SCAS scores in each story type condition, *F*(2, 78) = 0.45, *p* = .630, η_p^2 = 0.01. Scores were similar for the beach (*M* = 48.77, *SE* = 1.77), coming home (*M* = 51.08, *SE* = 1.79), and the park story (*M* = 50.44, *SE* = 1.84). The interaction between gender and age was also non-significant, *F*(2, 78) = 0.04, *p* = .961, η_p^2 = .01. Age was significantly correlated with SCAS scores, *r* = −.26, *p* = .019, 95% BCa CI [-0.50, −0.02], indicating that trait anxiety decreased significantly with age.

Emotion Ratings

Happiness. There was a significant main effect of the valence of story on happiness ratings, *F*(2, 75) = 7.44, *p* = .001, η_p^2 = 0.17. *Bonferroni post hoc* tests revealed that children reported significantly greater levels of happiness in the positive condition (*M* = 3.27, *SE* = 0.18) compared to the negative condition (*M* = 2.35, *SE* = 0.18), *p* = .004, 95% BCa CI [0.31, 1.53], *d* = 0.88, and in the neutral condition (*M* = 3.08, *SE* = 0.18) compared to the negative condition, *p* = .003, 95% BCa CI [0.11, 1.35], *d* = 0.85. Although greater levels of happiness were reported in the positive compared to the neutral condition, this difference was not significant, *p* = .419, 95% BCa CI [−0.30, 0.65], *d* = 0.23. There was no significant effect of story type, *F*(2, 75) = 0.30, *p* = .740, η_p^2 = 0.01 or interaction between story type and valence, *F*(4, 75) = 0.99, *p* = .419, η_p^2 =.05.

Internalization. There was a significant main effect of the valence of story on internalization ratings (fear and sadness combined), $F(2, 75) = 13.90$, $p < .001$, $\eta_p^2 = .27$. *Bonferroni post hoc* tests revealed that children reported significantly greater levels of internalization in the negative condition ($M = 1.15$, $SE = 0.14$) compared to the neutral condition ($M = 0.26$, $SE = 0.14$), $p = .001$, 95% BCa CI [0.52, 1.23], $d = 1.22$, and compared to the positive condition ($M = 0.27$, $SE = 0.14$) $p = .001$, 95% BCa CI [0.48, 1.24], $d = 1.14$. The difference between internalization ratings in the neutral and positive conditions was not significant, $p = .948$, 95% BCa CI [−0.34, 0.34], $d = 0.03$. The main effect of story type was not significant, $F(2, 75) = 0.14$, $p = .871$, $\eta_p^2 = .00$, nor was the interaction between story type and valence, $F(4, 75) = 1.10$, $p = .363$, $\eta_p^2 = .06$.

Properties of DRM Lists

A series of *Bonferroni* adjusted ANOVAs was run to test whether words lists differed across each property. Valence and arousal norms were taken from the Warriner, Kuperman, and Brysbaert norms for 13,915 English lemmas (2012). Most research uses norms from Bradley and Lang’s (1999) Affective Norms for English Words (ANEW). However, this contains only 1,034 words, including 12 from the negative, 8 from the neutral, and 13 from the positive lists. In comparison, the Warriner at el norms included 27 of the negative, and 27 of the neutral, and all 28 of the positive words presented. Ratings could range from 9 (happy; excited) to 1 (unhappy; calm). Mean valence and arousal scores for each list, and each critical lure (unpresented list name) are shown in Table 1, in addition to combined scores for each valence.

Table 1

Mean valence and arousal ratings for presented word lists and unpresented critical lures.

Valence List	*N*	Valence: List *M* (*SD*)	Valence: Critical Lure *M* (*SD*)	Arousal: List *M* (*SD*)	Arousal: Critical Lure *M* (*SD*)
Negative	*27*	*3.28 (1.24)*		*4.90 (1.08)*	
Thief	7	2.79 (0.60)	2.32 (1.34)	5.32 (1.11)	6.05 (2.38)
Anger	7	2.78 (0.69)	2.50 (1.36)	5.86 (0.64)	5.93 (2.77)
Sick	6	3.81 (1.71)	2.29 (1.38)	4.49 (0.53)	4.67 (2.58)
Cry	7	3.83 (1.50)	3.22 (2.41)	3.88 (0.73)	5.45 (2.82)
Neutral	*27*	*5.90 (0.66)*		*3.55 (0.71)*	
Foot	6	5.18 (0.48)	4.68 (1.42)	3.73 (0.99)	2.77 (1.57)
Chair	7	5.82 (0.45)	5.88 (1.33)	3.00 (0.36)	2.86 (2.01)
Green	7	6.07 (0.67)	6.29 (2.33)	3.61 (0.70)	4.07 (2.59)
Fruit	7	6.41 (0.44)	7.00 (1.97)	3.87 (0.51)	4.09 (2.49)
Positive	*28*	*6.84 (0.79)*		*4.20 (0.96)*	
Music	7	6.42 (0.42)	7.67 (1.82)	4.17 (0.68)	5.57 (3.09)
Soft	7	6.43 (0.71)	7.13 (1.66)	3.24 (0.38)	3.04 (2.46)
Love	7	7.68 (0.30)	8.00 (1.39)	4.81 (1.18)	5.36 (3.23)
Sweet	7	6.84 (0.90)	7.77 (1.38)	4.57 (0.65)	4.14 (2.92)

Note: The words ‘sandals’ from the ‘Foot’ list, and ‘ill’ from the ‘Sick’ lists were not included in the database. The word ‘colour’ from the ‘Green’ list was listed as ‘color’, and the word ‘tears’ from the ‘Cry’ list was listed as ‘tear’.

Positive lists (*M* = 6.87, *SE* = 0.18) were rated as significantly more positive than neutral words on average (*M* = 5.90, *SE* = 0.18), *p* = .001, 95% BCa CI [0.57, 1.34], *d* = 1.30, which were rated as significantly more positive than negative words on average (*M* = 3.28, *SE* = 0.18), *p* < .001, 95% BCa CI [2.05, 3.18], *d* = 2.62, *F*(2, 79) = 107.00, *p* < .001, η_p^2 =.73. There was also a significant difference in arousal scores, *F*(2, 79) = 14.27, *p* < .001, η_p^2 =.27. Negative words (*M* = 4.90, *SE* = 0.18) were rated significantly higher on arousal than positive (*M* = 4.20, *SE* = 0.18) *p* = .019, 95% BCa CI [0.09, 1.32], *d* = 0.69, which were in turn rated significantly higher on average than neutral words (*M* = 3.55, *SE* = 0.18), *p* = .034, 95% BCa CI [0.04, 1.26], *d* = 0.77.

Norms and values for a number of other properties of words were acquired from N-Watch; a program for deriving neighbourhood size and other psycholinguistic statistics (Davis, 2005). These properties can affect reaction times, lexical decision making, word identification, perceptual identification and semantic categorisation; potential confounds in memory recognition tasks (Atkinson & Juola, 1971; Brown & Watson, 1987; Carroll & White, 1973; James, 1975; Luce & Pisoni, 1998; Morrison & Ellis, 1995; Whaley, 1978).

Overall average word frequency (per million words), mean word length (characters per word), spoken, written, and subjective word frequency, phoneme and syllable length, neighbourhood size, subset and superset frequency scores did not differ significantly between word lists.[19]

However, average subjective ratings of familiarity did differ between lists significantly, $F(2, 65) = 3.51$, $p = .036$, $\eta_p^2 = .10$. Familiarity was measured on a scale of 100 (very unfamiliar) to 700 (very familiar). These were scaled up from Likert scale responses of '1' to '7'. Pairwise comparisons revealed that neutral words ($M = 563.64$, $SE = 9.20$, 95% BCa CI [545.00, 581.93]) were rated as significantly more familiar, on average, than negative words ($M = 529.10$, $SE = 10.03$, 95% BCa CI [504.63, 553.53]), $p = .022$, $d = 0.71$. However, their confidence intervals overlap, suggesting their true means could be of a similar value. There was no significant difference between the positive ($M = 556.91$, $SE = 9.80$, 95% BCa CI [540.65, 571.57]) and negative words, $p = .625$, $d = 0.16$, and positive and neutral words were not rated as significantly different, $p = .064$, $d = 0.60$.

Average age of acquisition differed significantly between the three lists, $F(2, 47) = 6.14$, $p = .005$, $\eta_p^2 = .21$. Scores also ranged from 100 to 700, and had been scaled up from responses of '1' to '7' representing years of age at acquisition. Pairwise comparisons revealed that the neutral words ($M = 253.79$, $SE = 13.64$, 95% BCa CI

---

19 Full means and ANOVA statistics for each property are reported in the Appendix, along with further detail on corpus databases.

[229.14, 282.90]) were rated as having an earlier mean age of acquisition than the negative words (*M* = 323.38, *SE* = 14.87, 95% BCa CI [293.58, 354.92]), *p* = .002, *d* = 1.19. The positive words (*M* = 274.07, *SE* = 15.36, 95% BCa CI [243.11, 307.49]) also had an earlier age of acquisition on average than negative words, *p* = .040, *d* = 0.79, but with overlapping confidence intervals around their means. There was no significant difference between positive and neutral words, *p* = .352, *d* = 0.35.

Average imageability ratings also differed significantly between the three lists, *F*(2, 65) = 3.89, *p* = .025, η_p^2 =.11. Scores could range from 100 to 700, with a higher score indicating greater ease in forming a mental image from the word. Pairwise comparisons revealed that the neutral words (*M* = 581.60, *SE* = 14.21, 95% BCa CI [557.88, 604.38]) had higher average ratings of imageability than the negative words (*M* = 523.00, *SE* = 15.50, 95% BCa CI [488.08, 555.85]), *p* = .006, *d* = 0.87. The positive words (*M* = 556.73, *SE* = 15.14, 95% BCa CI [520.94, 589.54]) were not scored significantly differently to the negative, *p* = .164, *d* = 0.43, or neutral words, *p* = .228, *d* = 0.36.

Recognition Task

Children's hits and false alarms across lists of the same valence were summed, and converted into hit and false alarm rates. The method outlined in Chapter I was used to calculate *d*' and *C* scores. Two participants were excluded from recognition analysis because at least one of their *d*' scores were less than 3 *SD*s below the mean. Table 2 shows the hit rates and false alarm rates across each condition. False alarm rates are broken down for critical lures, related distractors and unrelated distractors.

Table 2

Mean proportion of hits and false alarms by valence.

	Positive		Neutral		Negative	
Response rates	*M*	*SE*	*M*	*SE*	*M*	*SE*
Hits	0.89	0.01	0.83	0.01	0.86	0.01
False alarms	0.11	0.01	0.08	0.01	0.10	0.01
Critical lures	0.22	0.03	0.26	0.03	0.17	0.03
Related distractors	0.10	0.02	0.09	0.02	0.28	0.03
Unrelated distractors	0.08	0.01	0.05	0.01	0.05	0.01

Sensitivity: *d*' analysis. There was a high level of discrimination overall, with one sample *t*-tests showing that participants could discriminate at better-than-chance levels in all conditions, with all *d*' scores significantly greater than 0. For the positive lists $t(81) = 33.11$, $p < .001$, 95% BCa CI [2.29, 2.58], $d = 7.36$, neutral lists $t(81) = 31.94$, $p < .001$, 95% BCa CI [2.40, 2.74], $d = 7.10$, and for negative lists $t(81) = 33.66$, $p < .001$, 95% BCa CI [2.47, 2.81], $d = 7.48$.

A 3 × 3 × 6 (Story valence [positive, neutral, negative], DRM list valence [positive, neutral, negative], Rotation order [1, 2, 3, 4, 5, 6]) mixed ANOVA was run. [20] The main effects of rotation order, $F(5, 64) = 1.79$, $p = .127$, $\eta_p^2 = .12$, and story valence, $F(2, 64) = 1.68$, $p = .194$, $\eta_p^2 = .05$, on *d*' scores were non-significant. The rotation × story valence interaction was also non-significant, $F(10, 64) = 0.70$, $p = .722$, $\eta_p^2 = .10$.

There was a significant main effect of list valence on *d*' scores, $F(2, 128) = 4.31$, $p = .015$, $\eta_p^2 = .06$. Children could discriminate negative signals from noise ($M = 2.65$, $SE = 0.08$) significantly better than positive signals from noise ($M = 2.44$, $SE = 0.07$), $p = .010$, 95% CI [−0.15, 0.31], $d = 0.23$. There was no significant difference in

[20] Analysis was run with age, gender and trait anxiety scores entered as covariates. None were found to significantly predict *d*' scores or interact significantly with independent variables, so were not retained in the final analysis. Results from the covariate analysis can be found in the supplementary material.

discriminability between negative and neutral signals (*M* = 2.58, *SE* = 0.08), *p* = .173, 95% CI [−0.51, 0.43], *d* = 0.08, or between neutral and positive signals, *p* = 1.000, 95% CI [−0.31, 0.15], *d* = 0.16. However, the interaction between rotation order and DRM list valence was significant, *F*(10, 128) = 2.83, *p* = .003, η_p^2 = .18. A series of repeated measures ANOVAs revealed that for those who received the lists in the negative-positive-neutral order, there was a significant difference between *d*' scores depending on list valence, *F*(2, 28) = 4.11, *p* = .027, η_p^2 = .23 (see Figure 1). Discrimination scores were significantly greater for negative items (*M* = 3.09, *SE* = 0.15) than for positive items (*M* = 2.39, *SE* = 0.17), *p* = .003, 95% CI [0.25, 1.16], *d* = 1.09. Neutral *d*' scores (*M* = 2.52, *SE* = 0.20) did not differ significantly from negative, *p* = .194, 95% CI [−1.35, 0.20], *d* = 0.27, or positive *d*' scores, *p* = 1.000, 95% CI [−0.71, 0.98], *d* = 0.11. There was also a significant effect of valence on *d*' scores for those in the negative-neutral-positive order of presentation, *F*(2, 20) = 5.93, *p* = .010, η_p^2 = .37, again with significantly greater scores for the negative items (*M* = 2.74, *SE* = 0.24) compared to the positive items (*M* = 1.82, *SE* = 0.21), *p* = .021, 95% CI [0.14, 1.70], *d* = 1.04, and no significant difference between neutral *d*' scores (*M* = 2.35, *SE* = 0.29) and negative *d*' scores, *p* = .447, 95% CI [−1.11, 0.33], *d* = 0.48, or positive *d*' scores, *p* = .265, 95% CI [−0.27, 1.32], *d* = 0.59. There were no significant differences by valence within the positive-neutral-negative, *F*(2, 28) = 1.55, *p* = .231, η_p^2 = .10, positive-negative-neutral, *F*(2, 26) = 0.35, *p* = .711, η_p^2 = .03, neutral-positive-negative, *F*(2, 24) = 1.13, *p* = .340, η_p^2 = .09, or neutral-negative-positive order of presentation, *F*(2, 26) = 0.65, *p* = .529, η_p^2 = .05.

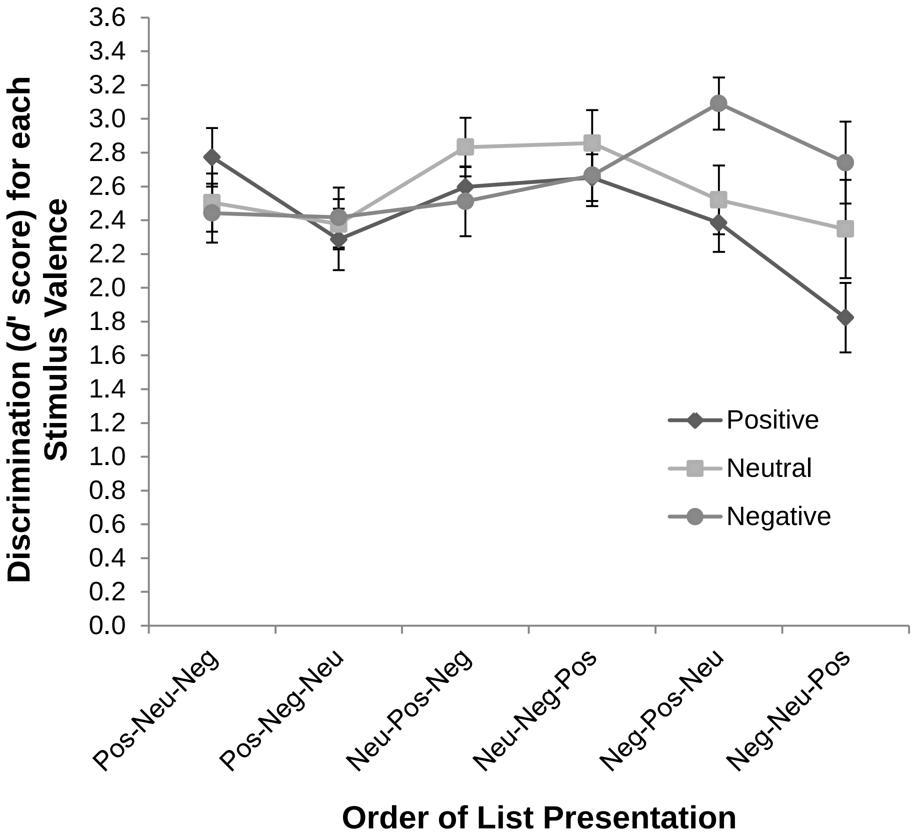


Figure 1. Line chart showing *d*' scores (discrimination) for positive, neutral and negative stimuli by order of presentation. Bars show standard error.

The list valence × story valence interaction was not significant, $F(4, 128) = 0.85$, $p = .494$, $\eta_p^2 = .03$, nor was the three-way interaction between presentation order, list valence and story valence, $F(20, 128) = 0.98$, $p = .488$, $\eta_p^2 = .13$.

Response bias: Criterion *C* analysis. *C* scores were significantly greater than 0 in the positive list condition, $t(81) = 4.70$, $p = .001$, 95% BCa CI [0.10, 0.23], $d = 1.04$, and neutral list condition, $t(81) = 6.94$, $p = .001$, 95% BCa CI [0.17, 0.32], $d = 1.54$. This indicates a conservative response bias, meaning that children were more prone to responding ‘false’ over ‘true’. They would therefore miss more items that had been presented, compared to falsely recognizing items that had not been presented. However, in the negative list condition, *C* was not significantly different to 0, $t(81) = 2.01$, $p =$

.053, 95% BCa CI [−0.00, 0.19], *d* = 0.45. This indicates no significant bias in responding with 'yes' or 'no'.

As with *d*' scores, a 3 × 3 × 6 (Story valence [positive, neutral, negative], DRM list valence [positive, neutral, negative], Rotation order [1, 2, 3, 4, 5, 6]) mixed ANOVA using *Bonferroni* corrections was conducted with *C* scores as the dependent variable.[21] There was a significant difference in *C* scores by list valence, *F*(2, 128) = 7.35, *p* = .001, η_p^2 = .10. Neutral *C* scores (*M* = 0.22, *SE* = 0.03) were significantly higher than negative *C* scores (*M* = 0.07, *SE* = 0.05), *p* = .003, 95% CI [−0.27, −0.04], *d* = 0.38, indicating greater conservative bias for neutral than negative list items. There was no significant difference between negative and positive (*M* = 0.17, *SE* = 0.04), *p* = .056, 95% CI [−0.20, 0.00], *d* = 0.20, or neutral and positive *C* scores, *p* = .375, 95% CI [−0.03, 0.14], *d* = 0.23.

However, the interaction between list valence and rotation order was also significant, *F*(10, 128) = 3.12, *p* = .001, η_p^2 = .20 (see Figure 2). For the positive-neutral-negative order of presentation there was a significant effect of list valence on *C* scores, *F*(2, 28) = 5.73, *p* = .008, η_p^2 = .29, with significantly greater conservative bias towards the neutral (*M* = 0.37, *SE* = 0.08) than the positive items (*M* = 0.13, *SE* = 0.06), *p* = .005, 95% CI [0.07, 0.42], *d* = 1.03. The differences in *C* scores between negative (*M* = 0.36, *SE* = 0.08) and positive, *p* = .063, 95% CI [−0.01, 0.48], *d* = 0.68, and between negative and neutral items, *p* = 1.000, 95% CI [−0.25, 0.23], *d* = 0.03, were not significant. There was a significant effect of list valence on *C* in the negative-neutral-positive condition, *F*(2, 20) = 7.78, *p* = .003, η_p^2 = .43 with significantly greater conservative bias towards the neutral (*M* = 0.13, *SE* = 0.12), and liberal bias towards the

[21] As with discrimination scores, age, gender and trait anxiety scores were entered as covariates and not found to significantly predict *C* scores or significant interact with independent variables. Results of this analysis can be found in the supplementary materials.

negative items ($M = -0.26$, $SE = 0.12$), $p = .010$, 95% CI [0.10, 0.70], $d = 1.15$, but no significant differences between the negative and positive ($M = 0.02$, $SE = 0.12$), $p = .064$, 95% CI [−0.58, 0.02], $d = 0.82$, or positive and neutral items, $p = .873$, 95% CI [−0.41, 0.18], $d = 0.34$. There was also a significant effect of list valence on *C* scores in the negative-positive-neutral condition, $F(2, 28) = 8.42$, $p = .001$, $\eta_p^2 = .38$ with significantly greater conservative bias towards positive ($M = 0.32$, $SE = 0.05$) and neutral ($M = 0.43$, $SE = 0.06$), compared to negative items ($M = 0.06$, $SE = 0.09$), $p = .036$, 95% CI [0.02, 0.49], $d = 0.82$; and $p = .021$, 95% CI [0.05, 0.69], $d = 1.44$, respectively. The difference between positive and neutral items was not significant, $p = .256$, 95% CI [−0.29, 0.06], $d = 0.65$.

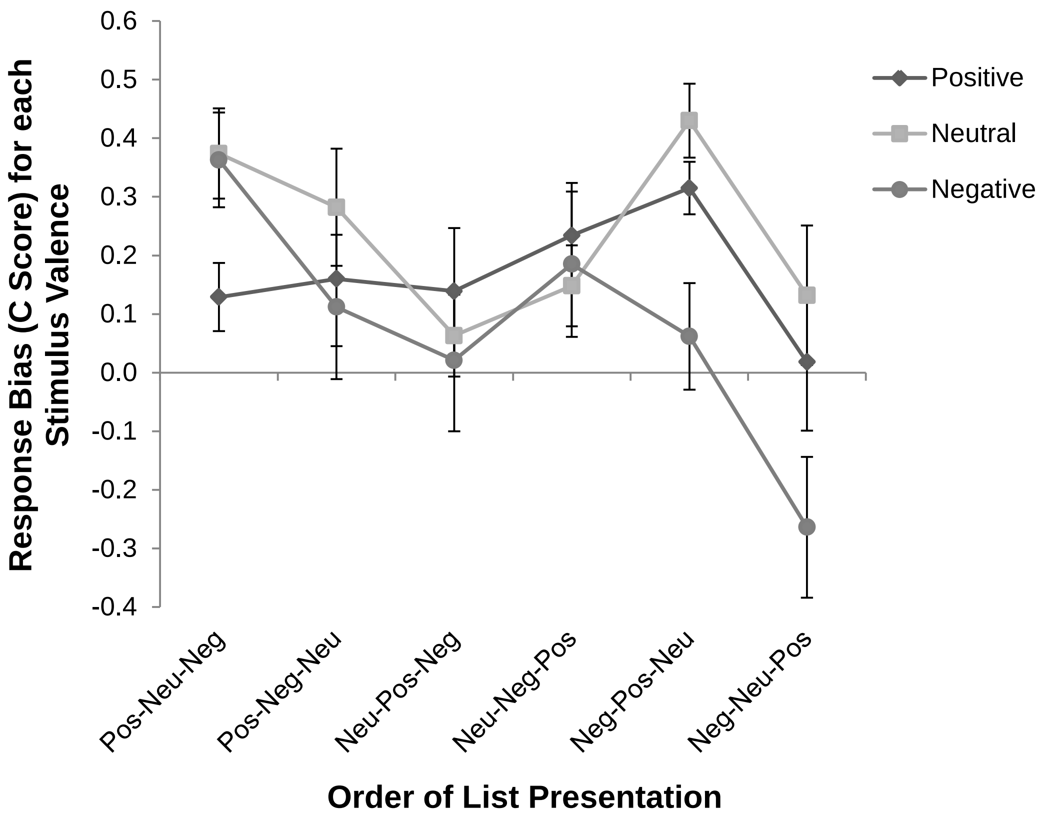


Figure 2. Line chart showing response bias (*C* scores) for positive, neutral, and negative stimuli by order of presentation. Bars show standard error.

List valence did not significantly affect *C* scores in the positive-negative-neutral, $F(2, 26) = 1.09$, $p = .351$, $\eta_p^2 = .08$, neutral-positive-negative, $F(2, 24) = 0.70$, $p = .508$, $\eta_p^2 = .06$, or the neutral-negative-positive condition, $F(2, 26) = 0.55$, $p = .586$, $\eta_p^2 = .04$. There was no significant main effect of rotation order, $F(5, 64) = 2.30$, $p = .056$, $\eta_p^2 = .15$, or story valence, $F(2, 64) = 0.41$, $p = .667$, $\eta_p^2 = .01$ on *C* scores. The rotation × story valence interaction was non-significant, $F(10, 64) = 1.06$, $p = .404$, $\eta_p^2 = .14$, as was the story valence × list valence interaction, $F(4, 128) = 0.12$, $p = .974$, $\eta_p^2 = .00$, and the three-way interaction between story valence, list valence, and rotation, $F(20, 128) = 1.09$, $p = .372$, $\eta_p^2 = .15$.

Discussion

This experiment had several aims: (1) to explore negativity and positivity preferences in an immediate recognition task; (2) to explore context effects between mood (induced by a valenced story) and discrimination for valenced word lists; and (3) explore potential moderators of emotional memory in children (age, gender, trait anxiety).

It was hypothesised that: (1) negative words would be discriminated at a lower rate than neutral words; (2) words matching the valence of the story would be discriminated at a higher rate than those that do not; (3) discrimination would increase with age; and (4) there would be no gender effect on discrimination.

There was no subsequent main effect of mood on recognition rates. There was a large effect of valence of word list, with negative words discriminated significantly more successfully than positive words, but only when they were presented first. This was mirrored by response bias, where lower levels of conservative response bias were observed for negative than positive and neutral words, when negative words were presented first. Lower levels of conservative bias were also observed for positive words when positive words were presented first, and negative presented last. There were no apparent context or moderator effects on discrimination.

Effects of Mood on Memory for Valenced Words

Those in the negative story condition reported significantly lower levels of happiness, and higher levels of internalizing ratings than those in the neutral and positive story conditions, whose ratings did not differ significantly. Because the manipulation of story valence was only partially successful, it is not possible to say with certainty whether there was genuine lack of mood priming effect, or whether the positive stories were not sufficiently emotive to bring out a strong enough effect on memory to reach significance. The negative stories evoked significantly lower levels of happiness (medium effect) and higher levels of internalization (large effect) than the positive and neutral stories. The difference in happiness ratings between the positive and neutral stories did not reach significance and was slightly below the cut off for a small effect (Morris & Fritz, 2013), though nevertheless, was in the right direction.

The stories had previously been used in an experiment by Bishop et al. (2004) as valenced stimuli, where all three had been rated as differing significantly on happiness in the expected directions. However, the stories were rated by a small number of adults rather than children, who may report that different materials evoke greater happiness, compared to adults.

Valence Effects of DRM List on Memory

Overall, children could discriminate negative signals from noise significantly better than positive signals from noise. However, there were no significant differences in discriminability between negative and neutral signals, or between positive and neutral signals. These findings tentatively suggest no negativity or positivity preference in recognition memory for word lists, in opposition to hypothesis (1). The valence effect between negative and positive words was significant and large, in both conditions where the negative words were presented first, suggesting genuine effects. In addition, this valence effect in discrimination was mirrored by a valence effect in response bias analysis; where less conservative bias was observed for negative, than positive and neutral words, when negative words were presented first. This means children were both more easily able to discriminate signals from noise, and less biased in responding according to some other criterion for negative words. Additionally, there was less conservative bias for positive words than neutral, in the condition where positive words were presented before neutral and then negative words. This means that in three of the four conditions where emotional words were presented first, children were less likely to ‘miss’ items for the first block. This advantage was not evident for neutral words, when they were presented as the first block.

Only a small number of previous experiments compare recognition of positive and negative words using a DRM paradigm in non-clinical samples of children, with mixed findings. Results of the current experiment do not fit with previous findings, where positive words were recognized more accurately than negative (Brainerd et al., 2010; Moradi et al., 2000), or no difference was observed (Neshat-Doost et al., 1998; Quas et al., 2015).

However, a stronger negativity than positivity preference is shown in recognition tasks in adults (Murphy & Isaacwitz, 2008). Neurological evidence suggests that explicit memory for both pleasant and unpleasant emotional stimuli can be enhanced through modulation of encoding and consolidation processes. Hamann (2001) reviewed evidence of neuroimaging studies, finding reports of increased amygdala activity when emotional film, positively and negatively valenced images, and negative photographs were presented, compared to neutral counterparts. Converging evidence is also presented from neuropsychological studies, whereby emotional memory enhancement is no longer found after damage to the amygdala. However, when word list stimuli were used, there was still an increase in amygdala activity, but this did not correspond with recognition scores for negative words. Hamann postulates that there may be a minimum threshold of emotional arousal below which memory enhancement will not be observed.

Findings are in line with Experiment 2, where negative narrative (verbal) material was discriminated more accurately than positive. Although the current experiment employed a different type of stimuli, narratives were presented to children, prior to the word lists. Although care was taken to ensure that none of the words from the lists were included in the stories, it is possible that the word lists may have been processed differently simply by following the presentation of a more complex narrative stimulus. For example, they may have acted as a mood primer, thus increasing arousal for the first set of word seen, with this effect wearing off for subsequent word lists. This could account for the rotation effects observed.

Alternatively, Haviland and Clark (1974) theorize that in comprehending a sentence, individuals first search their memory for antecedent information that matches information in the current sentence. Memory is then revised by attaching the new information. Their experiment showed that including a preceding sentence of the same context increased speed of comprehension. In the current experiment, the preceding story did not provide contextual semantic information, and it is possible that comprehension of the words would therefore be slower whilst recent memory was being unsuccessfully searched. Previous research did not include any contextual information, such as a narrative prior to presenting word lists (Brainerd et al., 2010; Moradi et al., 2000; Neshat-Doost et al., 1998; Quas et al., 2015).

Previous DRM experiments with neutral and negative word lists have consistently found a preference away from negative words relative to neutral words, with lower hit rates and higher false alarm rates (Howe, 2007; Howe et al., 2010, Exp 2, 4, and 5) and lower discrimination for negative words (Moradi et al., 2000; Neshat-Doost et al., 1998). This difference in findings could be due to methodological differences between the current and previous experiments. In all of Howe’s studies, a distraction task was administered followed by a recall task after each word list. When all lists had been presented, the recognition task was administered with emotional and neutral words presented randomly within one block. Similarly, in the other two studies (Moradi et al., 2000; Neshat-Doost et al., 1998) all words were presented in one list prior to a single distraction task, recall task, and recognition task with one mixed block of emotional and neutral words. The current experiment did not integrate a recall or distraction task, and the test phase was conducted separately for each block of word lists. The addition of the recall task previously could have impacted performance on the subsequent recognition task. Previous research shows children were less likely to report as many details after an emotionally negative experience when asked open ended questions, compared to more direct questions (Quas et al., 1999). The recall task in both studies could be conceptualised as an open-ended question, meaning that children are less likely to report as many items from the negative condition than neutral condition. The recall task would then act as an extra step of rehearsal, with increased rehearsal shown to systematically increase recognition rates (Woodward Jr., Bjork, & Jongeward Jr., 1973). As more words would likely be rehearsed in the neutral condition via recall, this could have led to the higher recognition rates of neutral words in the subsequent recognition task.

Even fewer studies analyse or report response bias in emotional memory studies; however, Brainerd et al. (2010) found that bias scores were lower (more conservative) for negative than positive words. Again, this was the reverse of the pattern observed in the current experiment. This is to be expected, given the reverse findings for discrimination.

Both confounding variables and power must be considered when interpreting the discrimination and response bias effects. As well as differing on valence, negatively valenced word lists were significantly more arousing than the positive lists, which in turn were significantly more arousing than the neutral words used in this study. Arousal positively predicts memory performance (Lang, Dhillon, & Dong, 1995), fitting with the higher rate of discrimination for negative lists observed in the current experiment, but not the comparison between positive and neutral words. Age of acquisition was significantly later for the negative than neutral and positive word lists, upon inspection of both the *p* value and confidence intervals. Age of acquisition has been show to positively predict recognition memory (Cortese, McCarty, & Schock, 2014), with this pattern observed in the current experiment. The word lists also differed significantly on ratings of imageability, with negative words scoring significantly lower than neutral words, but no different to positive words. Imageability positively predicts recognition memory (Cortese et al, 2014), meaning the neutral words should have been discriminated more accurately than the positive and negative words. However, this pattern was not observed. Cortese, Khanna, & Hacker (2010) found that age of acquisition was a much stronger predictor than imageability, which provides a possible explanation for discrimination results not matching that predicted by imageability scores. It is likely that the effects of valence, arousal, age of acquisition, and imageability interacted within the current experiment. Differences in arousal, and age of acquisition in particular could be driving the observed findings. Future experiments would need to ensure word lists were controlled on all characteristics except valence to test valence effects in isolation, successfully.

Sample sizes were too small to achieve a reasonable level of power once rotation order had been included as a fixed effect, in line with previous analysis designs used within this thesis. Other true effects may therefore have been missed, resulting in acceptance of the null hypothesis (a Type II error, or false negative). When only story and list valence were included, nearly 30 observations were observed per cell (26-28); once rotation was also included this greatly reduced power, with only 3-6 observations per cell. There are two alternative methods of analysis which could be considered; (1) removing story valence as an independent variable (due to the manipulation not being fully successful); or (2) removing rotation as a fixed effect (because it lowers power too much, and its inclusion was more exploratory, based on evidence from the previous two experiments reported in this thesis). The first approach is not generally advisable because it is considered bad practice to remove variables from analysis that were part of the original design. Guidelines for authors advise that all experimental conditions, including failed manipulations, be reported to ensure that false positives are not reported (Simmons et al., 2011). Removing the story valence (mood) condition would also not increase power substantially, resulting in an average of only 14 observations per cell. Additional analysis excluding rotation as a fixed effect found no significant mood, valence of word list, or context (interaction) effects.[22] Results of this experiment should be interpreted with caution, because the likelihood that a statistically significant result reflects a true effect is reduced in underpowered experiments (Button et al., 2013).

[22] Additional analysis is included in the appendix.

Models of emotional preferences propose that either all emotional material or just emotionally negative material is privileged in cognitive processing, compared to neutral material. Because no differences in discrimination between emotional and neutral words were identified in this experiment, support for either of these types of models is limited. Overall, discrimination was higher for negative than neutral words, but did not reach statistical significance. Despite this, a large effect was still observed. This effect is characteristic of a negativity preference. Taken with the large effect of valence between negative and positive words on discrimination, these two effects would support evolutionary models where threat related material is preferentially privileged, such as the evolved module of fear (Öhman & Mineka, 2001). However, on balance, when confounding variables differing between word lists, a lack of power, and previous experimental findings are considered, this support is considerably weaker, or indeed reversed.

It is not clear why order effects were observed within this experiment. Although rotation effects were also observed within Experiment 1 of this thesis, findings were mixed with geater liberal response bias for negative than positive visual items, but greater liberal response bias for the positive than negative narrative items when negative items were presented first; and more false memories recalled for the monster (negative) clip when the druid (positive) was presented first. However, no effects of rotation on discrimination were found in Experiment 1, and valence effects on discrimination do appear to vary substantially depending on stimulus type, making it hard to compare across experiments. These effects are inconsistent with previous research findings, which found no difference in discrimination depending on whether neutral or valenced lists were presented first (e.g. Howe, 2007; Howe et al., 2010).

Effects of Context on Recognition

Although the positive and neutral stories were not rated significantly differently to each other, the negative stories did evoke lower levels of happiness and higher levels of internalizing emotions than the others. Some of the expected effects could therefore still be tested; those who were presented with the negative story were still expected to show lower recognition than the other two groups, and to recognize the negative words more accurately than the positive and neutral.

The lack of any significant context effects is inconsistent with previous research (e.g. Bishop et al., 2004), and does not support hypothesis (2). One explanation for the failure to replicate these effects could be the study design. A repeated measures design was employed to allow children to participate in all three valence conditions of the recognition task. Bishop et al (2004) employed a between-participants design, whereby children where either in a low or high depressed group. In the current study, the valence of the first and second set of lists could have caused interference with any effect of context on subsequent lists. Additional analysis was therefore conducted on the effect of congruency of story valence and the valence of the first list on discrimination rates for the first block only, with the hypothesis that discrimination should be higher in the congruent condition (at least for negative words). However, there were still no significant effects of context or valence on recognition. [23] The lack of significant context effects can therefore not be explained by interference from multiple conditions.

The inconsistencies between previous and current findings could also be due to a focus on depression in previous studies, but anxiety in the current study. Mood-congruency memory effects have previously been found in non-clinical samples of children based on completion of the Depression Self Report Scale (Birleson, 1981).

---

[23] Additional analysis on effects of congruency of story valence and the valence of the first list presented on discrimination is included in supplementary material.

Williams et al. (1997) report that mood congruent memory preferences are reasonably well established in depression (but not in anxiety), whereas mood congruent attentional preferences are well established in anxiety (but not in depression). The current study employed a memory rather than attentional task, with the negative stories designed to evoke mild anxiety. It is therefore not surprising that a mood-congruency effect was not found. Although measures of anxiety and depressed mood (fear and sadness) were combined to include internalization as a moderator, neither measure predicted levels of discrimination nor interacted significantly with list valence, when two separate analyses were run. Alternatively, it is likely that the study was underpowered and may not have identified any true effects.

Moderators of Emotional Memory

Age, gender and trait anxiety were investigated as potential moderators of memory effects. With respect to age, the existing literature is unambiguous in the assertion that phonological short-term memory abilities develop rapidly through the early and middle childhood years (Gathercole, 1998). Previous research has also found an increase in hit rates for negative and neutral words with age (Howe, 2007; Howe et al, 2009, Howe et al; 2010) and for net accuracy and discrimination (Brainerd, 2010; Quas, 2013). Likewise, discrimination increased significantly with age in Experiment 1 of this thesis, and in the positive condition of Experiment 2. Support for hypothesis (3) was not found. Children in the current experiment showed higher levels of correct recognition and lower levels of false recognition than those in Howe’s studies, with discrimination reaching a near ceiling effect. This near-ceiling effect implies that the task may have been too easy, and did not allow any main effect of age to be borne out. Additionally, competing models predict either an increase, decrease or no change in negativity preference with age (Field & Lester, 2010). The lack of a significant interaction between the effects of age and stimulus valence on discrimination could mean that: (1) children are born with varying levels of fear, and associated processing preferences that remain stable (Field & Lester, 2010); (2) negativity bias develops prior to 7 years of age and remains fairly stable over time (Martin, Horder, & Jones, 1992); or (3) that any developmental changes are simply obscured due to the near ceiling effect.

The non-significant effects of gender are in-line with existing DRM experiments using word lists (e.g. Howe, 2007; Howe et al, 2009; Howe et al., 2010; Quas, 2015) and images (Cordon, 2013), along with Experiment 2 of this thesis, and support hypothesis (4). Although a main effect of gender was observed in Experiment 1, this was thought likely a spurious result.

Finally, trait anxiety did not significantly moderate discrimination scores on the recognition tasks. This finding is unsurprising given that previous studies reporting a negativity bias in anxious children are based on clinical samples, and featured attentional tasks (Dalgleish et al., 2003; Dalgleish et al., 2001). The current experiment specifically excluded children with a formal diagnosis of any emotional disorders, with trait anxiety scores mirroring those seen in the normal range in same-age samples, and included a memory task rather than attentional task. This finding is also in line with Experiment 2 of this thesis.

Conclusions

In relation to the experimental hypotheses: (1) the expected finding of lower discrimination for negatively valenced compared to neutral words was not observed; (2) there was no evidence that words matching the valence of a previously presented story were discriminated more easily; (3) there was no evidence of the expected increase in discrimination with age; and (4) support for a null effect of gender was observed.

The present study found that children could discriminate negative signals from noise more accurately than positive signals from noise, with this finding only significant when negative lists were presented first. There was also significantly less conservative bias for negative than positive and neutral words when negative words were presented first, implying that children were also using an alternative decision criterion. Results do not match findings from existing DRM studies using negative and neutral words with children. This could be due to the addition of the story prior to word presentation in the current experiment, or due to the confounding variables of increased arousal, and later age of acquisition for the negative words; both of which positively predict discrimination rates. Context effects were not borne out in the expected manner; this could be due to the partial failure to manipulate mood with the stories, or ceiling effects in the discrimination task. Additionally, effects of rotation were observed on discrimination and bias. The experiment was underpowered, and observed effects do not match previous findings. They are therefore considered spurious.

In future experiments, test materials should be piloted with the appropriate age range prior to running the experiment, to ensure they achieve the desired manipulation effect and are suitable for the age range. Any presentation stimuli such as word lists should be standardized on as many variables as possible, to limit confounds such as arousal effects. Methods of decreasing the ceiling effect should also be sought, such as increasing task difficulty, or including a distractor task, for example. It would also be of theoretical interest to explore potential mechanisms of valence effects on memory, for example by manipulating explicit elaboration.

Chapter V

Explicit Elaboration and Valence Effects on Children's Emotional Memory

Abstract

The roles of elaboration and valence on children's true and false memories were examined using the Deese-Roediger-McDermott paradigm. One-hundred and thirty 7-10 year old children were shown blocked negative-emotional or positive-emotional, and neutral word lists. Children rated words on ease of spelling (shallow encoding) or valence (deep encoding) during the presentation phase, then completed a distractor and recognition task. As in previous published experiments of a similar design, discrimination was significantly lower for negative than for neutral words. Discrimination was also lower for positive than neutral words, though not to a significant level. Significantly greater liberal bias was found for positive compared to neutral words, indicating a potentially underestimated measure of true sensitivity. Higher elaboration led to a greater propensity to respond positively, potentially reflecting greater activation of semantic networks. This did not interact with valence effects in the expected manner, but this is likely due to methodological differences to previous experiments and a lack of observed power in the current experiment. An interesting next step would be to replicate experiments with adults that have included unrelated and related valenced and neutral words, to explore the relative contributions of valence and semantic relatedness in children.

Explicit Elaboration and Valence Effects on Children's Emotional Memory

The reliability of children's memories is often questioned, particularly in relation to real-life emotional situations. Many classic studies have reported that children's memory is less accurate and more suggestible than adults when interviewed about a negative emotional event such as witnessing a kidnapping, being in a stressful medical setting or around sexual abuse (for a review see Bruck & Ceci, 1999). More recent studies report similar findings for less extreme emotional situations. For example, significantly more children recalled a false memory for a fictitious negative event than a fictitious neutral event within a school classroom setting, that they were questioned about (Otgaar, Candel, & Merckelbach, 2008).

However, both positive and negative preferences in memory, and other forms of cognitive processing, are evident across a variety of settings. This phenomenon is known as the positive-negative asymmetry (Lewicka et al., 1992). A positivity 'bias' was originally defined as the tendency to evaluate situations or stimuli more positively than they actually are (Czapiński, 1982). An affective negativity 'bias' was originally defined as negatively valenced stimuli having a larger impact on mood than positive stimuli of equal intensity (Lewicka et al., 1992). However, Murphy and Isaacowitz, (2008) noted the inconsistent use of the term '*bias*' within literature and more recently, the term '*preference*' has been adopted by some psychologists to show an increase in performance towards a valenced stimulus compared to a neutral baseline stimulus across a measurable domain.

This paper is primarily concerned with true and false memories of neutral, and positively and negatively valenced stimuli in children. An attentional preference towards negatively valenced material in both clinically anxious, and non-clinically anxious children has been shown in a meta-analysis (Bar-Haim et al., 2007). This preference has been characterised by an initial avoidance of attending to negative stimuli (Muris, Huijding, Mayer, Remmerswaal, & Vreden, 2009) followed by a subsequent delay in disengaging from the stimuli (McKenna & Sharma, 2004). Within memory, an emotional preference would refer to a tendency to report a greater proportion of negative or positive, compared to neutral material, or to increased hit or discrimination rates for negative or positive, compared to neutral material. It can also be applied to false alarm rates.

Some theoretical models propose that children are born with cognitive processing preferences towards negative stimuli. For example, Öhman and Mineka (2001) proposed a ‘fear module’ in mammals, preferentially activated by evolutionarily threatening stimuli such as spiders and snakes. Stimuli related to threats that are more difficult to survey, and movements that are difficult to predict, such as snakes and spiders generally have been shown to evoke alertness and wariness (Frijda, 1986). ‘Integral bias’ models propose that children are born with varying levels of cognitive preferences towards negative stimuli, which remain constant over time, whilst ‘moderation models’ of information processing are based on the assumption that children are born with a preference towards negative content and learn to moderate these biases to varying degrees as they develop with age (see Field & Lester, 2010; Hadwin & Field, 2010 for a review).

Other theories hypothesize that all evolutionarily advantageous (positively and negatively valenced) material is privileged (for example, Caccioppo et al., 1997; 1999; Hamann, 2001). The Network Affect Model proposes that valenced material is stored as central nodes with higher numbers of connections to surrounding nodes compared to other materials (Bower & Cohen, 1988), or that valenced material tends to prompt greater semantic elaboration, and be more organizationally cohesive in nature

(Kensinger, 2004; Maratos, Allan, & Rugg, 2000; Phelps et al., 1998; Talmi & Moscovitch, 2004). These increases in connection, elaboration, or cohesiveness are proposed to lead to a memory advantage for emotionally valenced items.

Despite these early findings and theories; findings from rigorously controlled laboratory experiments are limited and present inconsistent findings. The Deese–Roediger–McDermott (DRM) paradigm (Roediger & McDermott, 1995) has become a popular method of testing memory accuracy. The majority of experiments using word lists as stimuli have found a preference away from negative words (Howe, 2007; Howe et al., 2010, Experiments 2, 4, and 5; Moradi et al., 2000; Neshat-Doost et al., 1998). In addition, children’s false alarm rates for negative-emotional material increased over time, whereas they remained constant for neutral items. This indicates that discrimination is likely to decrease at a faster rate for emotionally negative, than neutral materials over time. Fewer experiments include positive words, with two reporting a preference away from positive words (Moradi et al., 2000; Neshat-Doost et al., 1998).

Evidence of a preference toward emotionally valenced words in immediate or short-term memory tasks is sparse; A DRM study in Chapter IV of this thesis (Experiment 3) was conducted with 7-11 year old children who had emotionally positive, negative and neutral lists of words presented visually and aurally, followed by an immediate recognition task after each presentation phase. A negativity preference was observed, with higher discrimination for negatively valenced words, but only when they were presented before the neutral and positively valenced words. A conservative response bias was also found in the neutral and positive conditions; meaning children were more likely to ‘miss’ items of that valence in the recognition task. The only other experiment reporting an emotional preference, found a negativity preference for picture stimuli (Cordon et al., 2013).

There were some important methodological differences between those studies which did, and did not find a negativity preference in DRM recognition tasks. Howe's studies employed both a recall and recognition task, whereas those experiments finding a negativity bias employed only a recognition task. The recall task in Howe's studies was completed prior to the recognition task; children have been shown to exhibit a preference to report fewer details after an emotionally negative experience when asked open ended questions, compared to more direct questions (Quas et al., 1999). The recall task could have acted as an extra stage of consolidation, which would have affected performance in the subsequent recognition task. Words were presented aurally in Howe's studies, whereas pictures were used by Cordon et al. (2013) and words were presented both visually and aurally in Experiment 3 of this thesis. Experiment 2 (Chapter III) of this thesis showed an interaction between stimulus modality and valence on discrimination rates, with items from the aurally presented narrative recognized more accurately than items from the accompanying images, when either stimulus sets were negatively valenced. This indicates a potential difference in the processing of emotionally negative visual and aural stimuli. Lastly, Howe presented items in order of backwards associative strength (BAS), whereas the lists were presented in a randomized order in Experiment 3, and Cordon et al. (2013) used images which were not connected by critical lures. Presenting items in order of increasing BAS has been shown to increase the number of lures or false memories recognized (Arndt, 2013), which could potentially affect stimuli of different valences differently. In addition, near-ceiling effects were observed in Experiment 3 of this thesis, and the experiment was underpowered, which could have resulted in some effects not being borne out fully.

Processing models offer several possibilities for how and where emotional preferences in memory occur. According to a working memory model, audio targets are attended to and rehearsed by a phonological loop before being recognized as a word and stored in verbal working memory, and visual targets are stored within iconic memory before being moved into working memory; a process known as encoding (Baddeley, 2000; Baddeley, Eysenck, & Anderson, 2009). There may be differences in the memory storage itself, such as whether information is stored in short or long term memory, and the capacity for storage. Maintenance rehearsal is required to keep information in working (short term) memory, whereas elaborative rehearsal is necessary to move information into long term memory, due to its depth of processing (Craik & Lockhart, 1972). Alternatively, a negativity preference could be associated with the act of retrieving encoded information from memory via recognition or recall, the final stage in memory processing (Atkinson & Shiffrin, 1968).

Hamann (2001) presented a review of neuroimaging and neuropsychological evidence, suggesting that explicit memory for both pleasant and unpleasant emotional stimuli can be enhanced through modulation of encoding and consolidation processes. If the negativity preference in memory shown in some DRM recognition tasks (Cordon et al., 2013; Experiment 3, Chapter IV) is due to a higher level of elaboration for valenced words, then removing the natural advantage of elaboration for valenced words during encoding should diminish or extinguish any negativity or valence effects by offering the same level of elaboration to all three categories of words. In the current study, there will be a low and high elaboration condition. In the low elaboration condition, children will be asked to rate each word on how hard it is to spell, which does not require semantic elaboration. It is expected that a similar or slightly diminished negativity preference will be observed compared to Experiment 3 (Chapter IV) due to the introduction of an encoding task. In the high elaboration condition, children will be asked to rate how each word makes them feel. Evaluating stimuli in relation to the self produces organized and elaborate processing, resulting in increased memory. A meta-analysis shows that this reliably increases memory performance; this is known as the self-reference effect (Symons & Johnson, 1997). Elaboration for both positive and neutral words is expected to increase in the high elaboration condition substantially, diminishing the valence effect even further. These two tasks were chosen because they could both be measured using a Likert scale, and varied in the level of semantic elaboration required.

In an effort to reduce the ceiling effects observed in the previous chapter, a distraction task was introduced between each presentation and recognition phase of the task. This task consisted of simple addition and subtraction maths sums, designed to be used in a classroom setting for children aged 5 years and over. This activity should limit the possibility of rehearsal due to domain-general suppression of previous, distracting material, but not interfere with any effects of the experimental manipulation due to a different domain of skills being utilised (Craik, 2014). The task is also designed to be easy enough not to add stress to the task for any of the children. Any stress experienced between presentation and recognition phases could act as a mood inducer, and confound the effects of the experimental manipulation due to context effects (Matt et al., 1992; Ucros, 1989).

The upper limit on the age range has also been lowered to 10 years of age to ensure that most children utilise as much of the spelling rating scale as possible by finding some of the words harder to spell than others. Lowering the maximum age should help to reduce ceiling effects because memory performance increases with age during childhood (Gathercole, 1998).

This experiment aims to explore mechanisms underlying observed negativity effects identified in Experiments 2 and 3 (Chapters III and IV) by manipulating elaboration levels. It is hypothesized that: (1) as levels of elaboration increase, the valence effect will diminish; (2) discrimination will increase with age; and (3) there will be no gender effects. Trait anxiety will not be included as a potential moderator due to a lack of significant findings in previous experiments, and a lower focus on anxiety in the experiment overall.

Method

Participants

Five schools in East Sussex, West Sussex and Surrey, England participated in the study. One-hundred and forty-six opt-in consent forms were returned by parents of children in school years 3-5 (age 7-10 years). Nine children were excluded because their siblings had taken part, three did not provide assent, two were unavailable during the sessions, and one did not meet inclusion criteria due to an uncorrected hearing impairment. One-hundred and thirty children took part in sessions. Fifty-five children (42.30%) were boys and 75 (58.70%) were girls. Ages ranged from 7 years, 4 months to 10 years, 3 months (*M* = 8 years, 9 months, *SD* = 8.91 months). All spoke English as their primary language, and none had a formal diagnosis of behavioural, developmental, or educational disorders based on parental report. All children had normal, or corrected to normal hearing and vision. The group was ethnically, culturally and socio-economically mixed, including pupils from an Independent Girls' School, three mixed Community State Schools, and a mixed voluntary aided Church of England primary school. Four of the five schools were smaller than average, with the remaining being much larger than average. All schools had a higher than average proportion of White British students, but report bilingual pupils and pupils of minority ethnicities. Four of the schools had a lower than average proportion of pupils eligible for pupil premium, with the remaining school reporting a higher than average proportion. One school was rated as ‘requires improvement’, three ‘good’ and one ‘excellent’, in reports by Ofsted and ISI. Children took part voluntarily and were naïve to the purpose of the experiment. Schools and children were entered into separate voucher prize draws on completion.

Materials

Ratings were recorded for how Happy, Sad, Excited and Scared children had felt that day on a 4-item self-report questionnaire using a 5-point Likert Scale from ‘0’ (*Not at all*) to ‘4’ (*Very, very much*).

There were six DRM stimuli lists, each consisting of 14 items (see Appendix: Chapter IV, V – Stimuli). These were the same lists as used in Experiment 3 in this thesis. Items were taken from 12 previously established backwards associative strength (BAS) DRM stimuli lists, with items 2-8 from each list included in the experimental stimuli lists. List A and B consisted of positive items (A [‘Sweet’ and ‘Love’], B [‘Soft’ and ‘Music’]), C and D of neutral items (C [‘Green’ and ‘Chair’], D [‘Fruit’ and ‘Foot’]), and E and F of negative items (E [‘Cry’ and ‘Thief’], F [‘Angry’ and ‘Sick’]). The lists Sweet, Chair, Fruit, Foot, and Cry were sampled from lists used by Budson et al. (2006), The lists Love, Soft, Music, Green, Thief, Angry, and Sick were sampled from the Cornell/ Cortland emotion lists (Brainerd et al., 2010). There was no overlap between items across lists. The order of items was randomized within each of the 12 previously established lists, with items from each list pair presented alternately to make the final six lists. The order of item presentation in each list remained fixed for all children.

The presentation paradigm was programmed in Matlab. A fixation point was displayed in the centre of the screen for 5 s prior to the trial commencing. Each item was presented in large black font on a white background for 2 s. Contrast and brightness of the screen was standardized across children. As each word appeared, a recording of the word was played through headphones in a British female voice, at a standardized volume. Following each item, a screen then appeared with a question and horizontal ratings scale. In the low elaboration condition, children were asked to rate each word on ease of spelling on a horizontal slider scale. The scale included 5 references points from '*very easy*' (−400 from centre), to '*very hard*' (400 from centre). In the high elaboration condition children were asked to rate how each word made them feel from '*very bad*' (−400 from centre), to '*very good*' (400 from centre). Once the participant clicked on the scale a blank screen with a fixation cross in the centre appeared for 2 s prior to the next item. The programme was set not to respond to any clicks on the scale made during the first 2 s of the question appearing, to avoid erroneous clicking, and children selecting the same answer without any thought.

The recognition paradigm consisted of 20 items. Two items were the critical lures from the stimuli lists (not presented in the presentation paradigm), two were related lures (the first item from each original BAS DRM list, not presented in the presentation paradigm), 10 were unrelated distractors from alternate DRM lists of the same valence, and six were target items (items in position 3, 5 and 7 from each original BAS DRM last presented in the presentation paradigm). The order of items was randomized but remained fixed between children. The Matlab recognition paradigm was the same as the ratings/ presentation paradigm, but instead of rating each item, children were asked 'Have you seen this word before?' and were asked to click inside a '*yes*' or '*no*' box as a forced choice answer.

Design

A mixed design was employed. Type of word was a within-participants variable with two levels (neutral vs valenced), whilst Valence of emotive list, with two levels (positive vs negative) and Elaboration, with two levels (low vs high) were between-participants variables. Children were assigned systematically to the four conditions; positive-low elaboration (*n* = 33), negative-low elaboration (*n* = 32), positive-high elaboration (*n* = 34) and negative-high elaboration (*n* = 32). The order of presentation was counterbalanced; half of the children in each condition were presented with the neutral lists first, half the valenced. Presentation order was included as a fixed effect in analyses.

Within the ratings task, the dependent variable was difficulty in spelling (from 'very easy' to 'very difficult') for the low elaboration group, and strength of emotion elicited (from 'very bad' to 'very good') in the high elaboration groups. The dependent variables in the recognition tasks were hit and false alarm rates, used to calculate *d*' prime and criterion *C* scores. For analysis of emotion and spelling ratings of words, *d*' scores, and *C* scores, a series of mixed ANCOVAs with *Bonferroni* adjustments was conducted. Age was included as a covariate in the first step and gender in the second. Covariates were retained only where they significantly predicted the outcome variable or interacted significantly with other independent variables.

Procedure

Interviews were conducted individually in a quiet space at the child's school, away from their classroom. Children had the interview procedure and confidentiality information outlined to them. They were given a chance to ask questions and assent was attained. For the self-report measures the researcher read the questions out loud, and the children circled their response on the answer sheet. Children were able to ask questions if they did not understand any of the items.

Each presentation paradigm was followed by one minute of simple maths sums as a distraction task, and then the matching recognition paradigm. This was repeated for the four word lists. Children were debriefed, thanked, awarded a sticker, and given the option of viewing the correct recognition answers before returning to class.

Children were monitored closely for any signs of distress, at which point the interview would be terminated immediately and children would have the choice of reading an age appropriate book before returning to class.

Results

Children's Mood Ratings

An independent samples *t*-test revealed no significant differences in initial ratings of happiness between children in the positive (*M* = 3.32, *SE* = 0.11) and negative conditions (*M* = 3.53, *SE* = 0.08), *t*(128) = −1.55, *p* = .123, 95% BCa CI [−0.49, 0.03], *d* = 0.27. A second independent samples *t*-test revealed no significant differences in initial ratings of internalization (combined fear and sadness scores) between children in the positive (*M* = 0.92, *SE* = 0.18) and negative conditions (*M* = 0.63, *SE* = 0.12), *t*(110.15) = 1.38, *p* = .123, 95% BCa CI [−0.09, 0.69], *d* = 0.24.

Properties of DRM lists

Properties of the DRM lists were analysed and reported in Chapter IV. Positive lists (*M* = 6.87, *SE* = 0.18) were rated as significantly more positive than neutral words on average (*M* = 5.90, *SE* = 0.18), *p* = .001, 95% BCa CI [0.57, 1.34], which were rated as significantly more positive than negative words on average (*M* = 3.28, *SE* = 0.18), *p* < .001, 95% BCa CI [2.05, 3.18], *F*(2, 79) = 107.00, *p* <.001, η_p^2 =.73. There were no significant differences in word frequency, word length, or neighbourhood size subset or superset frequency. However, there were significant differences in arousal scores, age of acquisition, and imageability between the word lists (*p*'s <.001, .005, .025, respectively). Full analysis is presented in Chapter IV and the Appendix (Chapter IV, V – Stimuli).

Ratings of Words in DRM Lists

Ease of spelling (low elaboration condition). Data from four children were excluded because their scores exceeded 3 *SD*s of the mean. Age was a significant predictor of spelling difficulty ratings, $F(1, 55) = 12.47$, $p = .001$, $\eta_p^2 = .19$, but did not interact significantly with word type, $F(1, 55) = 0.01$, $p = .902$, $\eta_p^2 = .00$. Spelling difficulty decreased as age increased for emotional ($\beta = -5.82$) and neutral words ($\beta = -5.92$). Gender scores did not significantly predict ease of spelling ratings ($p = .838$).[24]

A 2 × 2 × 2 (Affect condition [positive, negative]), Word type [emotional, neutral], Rotation [emotional first, neutral first]) mixed ANCOVA with *Bonferroni* adjustments was conducted on spelling difficulty ratings, with age entered as a covariate. Overall, children did not report significantly different ratings of spelling difficulty in the positive affect ($M = -248.69$, $SE = 21.37$) compared to the negative affect condition ($M = -205.25$, $SE = 20.66$), $F(1, 55) = 2.13$, $p = .150$, $\eta_p^2 = .04$. There was no significant difference reported in the difficulty of spelling for affective words ($M = -214.98$, $SE = 16.29$) compared to neutral words ($M = -238.96$, $SE = 14.47$), $F(1, 55) = 0.02$, $p = .893$, $\eta_p^2 = .00$. The interaction between the type of word and the affect condition on spelling ratings was not significant, $F(1, 55) = 1.09$, $p = .302$, $\eta_p^2 = .02$. The main effect of rotation on spelling ratings was not significant, $F(1, 55) = 0.77$, $p = .368$, $\eta_p^2 = .01$, and rotation did not interact significantly with word type, $F(1, 55) =$

---

24 Full covariate analysis results are shown in supplementary materials.

1.96, $p = .167$, $\eta_p^2 = .03$, valence condition $F(1, 55) = 3.46$, $p = .068$, $\eta_p^2 = .06$, or word type × valence, $F(1, 55) = 1.78$, $p = .188$, $\eta_p^2 = .03$.

Emotion evoked (high elaboration condition). Data from six children were excluded due to scores lying outside 3 *SD*s of the mean. Age and gender scores did not significantly predict ratings of word valence (*p*s = .977, .940, respectively).[25] A mixed 2 × 2 × 2 (Affect [positive, negative], Word type [emotional, neutral], Rotation [emotional first, neutral first]) ANOVA with *Bonferroni* adjustments showed that overall, significantly greater positive ratings were reported for neutral ($M = 72.96$, $SE = 8.32$) than for emotional words ($M = 21.57$, $SE = 9.91$), $F(1, 55) = 44.65$, $p < .001$, $\eta_p^2 = .45$. Children in the positive affect group rated words as eliciting significantly greater positive emotions ($M = 132.03$, $SE = 11.24$) than ratings in the negative affect condition ($M = -37.50$, $SE = 12.22$), $F(1, 55) = 104.24$, $p < .001$, $\eta_p^2 = .66$. There was also a significant interaction between the type of word, and affect condition on mean emotion ratings, $F(1, 55) = 301.12$, $p < .001$, $\eta_p^2 = .85$. Scores were highest (most positive) for the emotional words in the positive affect group ($M = 172.98$, $SE = 14.75$), and lowest (most negative) for the emotional words in the negative affect group ($M = -130.31$, $SE = 8.35$). The neutral words were rated as significantly more positive in the positive affect ($M = 90.35$, $SE = 13.38$) than in the negative affect condition ($M = 55.06$, $SE = 8.35$), $t(50.69) = 2.24$, $p = .026$, 95% BCa CI [8.09, 66.87.54], $d = 0.57$. Paired *t*-tests also showed significant differences in average ratings between neutral and emotional lists in the positive affect condition, $t(31) = 7.79$, $p < .001$, 95% BCa CI [62.32, 103.19], $d = 1.49$, and negative affect condition, $t(26) = -16.71$, $p < .001$, 95% BCa CI [−207.92, −163.84], $d = 3.36$.

25 Full covariate analysis results are shown in supplementary materials.

The main effect of rotation on valence ratings was not significant, $F(1, 55) = 0.31$, $p = .582$, $\eta_p^2 = .01$, and rotation did not interact significantly with word type, $F(1, 55) = 0.11$, $p = .743$, $\eta_p^2 = .00$, valence condition $F(1, 55) = 0.02$, $p = .877$, $\eta_p^2 = .00$, or word type × valence, $F(1, 55) = 2.28$, $p = .137$, $\eta_p^2 = .04$.

Recognition Task

Children's correct hits (H) and incorrect false alarms (F) across lists of the same valence were summed and converted into hit and false alarm rates. Discrimination *(d')* and criterion *C* scores were calculated for sensitivity and response bias analysis, as outlined in Chapter I. Data from three children were excluded from the recognition analysis due to missing trials resulting from software issues. Data from eight children were excluded from recognition analysis due to outlying *d*' or *C* scores (more than 3 *SD*s above or below the mean).

Sensitivity: Type I *d*' analysis. There was a high level of discrimination overall, with one sample *t*-tests showing that children could discriminate at better-than-chance levels for all word lists with *d*' scores significantly greater than 0. For the positive lists, $t(59) = 28.45$, $p < .001$, 95% BCa CI [2.52, 2.93], $d = 7.41$, neutral lists in the positive condition, $t(59) = 26.86$, $p < .001$, 95% BCa CI [2.64, 3.06], $d = 6.99$, neutral lists in the negative condition, $t(58) = 33.72$, $p = <.001$, 95% BCa CI [2.82, 3.18], $d = 8.80$, and negative lists $t(58) = 34.56$, $p < .001$, 95% BCa CI [2.37, 2.66], $d = 7.65$.

Discrimination (*d*') scores increased significantly with age, $F(1, 110) = 6.37$, $p = .013$, $\eta_p^2 = .06$, for neutral words, $\beta = .01$ and for emotional words, $\beta = .02$. There was no significant interaction between age and word type on *d*' scores, $F(1, 110) = 1.18$, $p = .278$, $\eta_p^2 = .01$. Gender did not significantly predict *d*' scores ($p = .958$).[26] A 2 × 2 × 2 × 2

[26] Full covariate analysis is shown in the supplementary materials.

(Word type [emotional, neutral], Affect [positive, negative], Elaboration [low, high], Rotation [emotional first, neutral first]) mixed ANCOVA using *Bonferroni* adjustments was conducted on *d*' scores, with age entered as a covariate.

Mean *d*' scores were not significantly different in the low (M = 2.75, SE = 0.08) and high elaboration condition (M = 2.81, SE = 0.08), after accounting for age, $F(1, 110) = 0.24$, $p = .623$, $\eta_p^2 = .00$, or in the positive (M = 2.79, SE = 0.08) compared to the negative affect condition (M = 2.77, SE = 0.08), $F(1, 110) = 0.04$, $p = .845$, $\eta_p^2 = .00$. The interaction effect between elaboration and affect condition on *d*' scores was not significant, $F(1, 110) = 1.21$, $p = .273$, $\eta_p^2 = .01$. There was no significant difference in *d'* scores between neutral (M = 2.93, SE = 0.07) and emotional words (M = 2.63, SE = 0.06), $F(1, 110) = 2.17$, $p = .143$, $\eta_p^2 = .02$. However, word type and word affect did interact significantly, $F(1, 110) = 8.03$, $p = .005$, $\eta_p^2 = .07$ (see Figure 1).

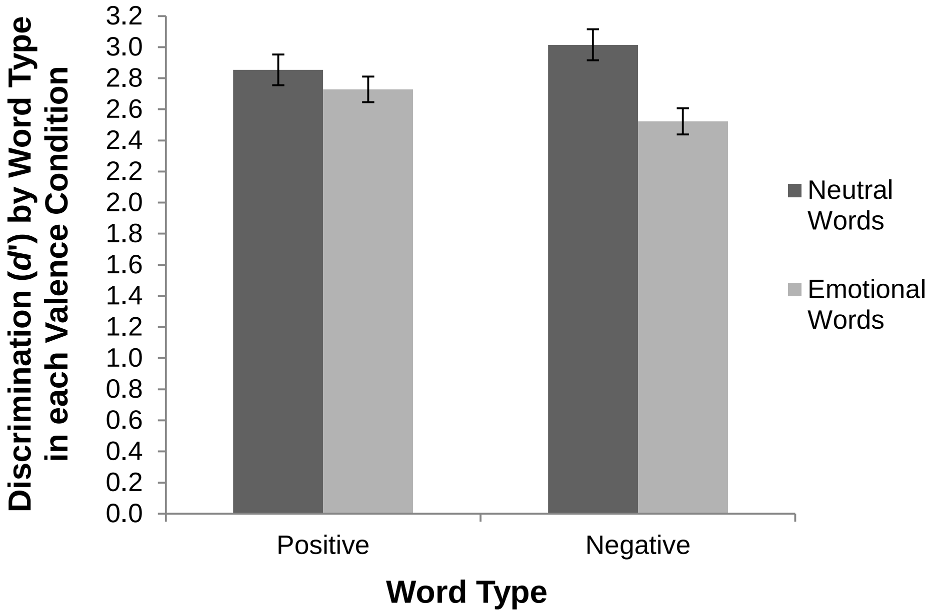


Figure 1. Clustered bar chart showing discrimination (*d*' scores) for emotional and neutral words in the positive and negative conditions. Bars show standard error.

Mean *d*' scores were higher for neutral words than emotional words in both conditions; mean *d*' scores for neutral words did not differ significantly between the negative condition (M = 3.00, SE = 0.10, 95% CI [2.81, 3.20]), and the positive condition, (M = 2.85, SE = 0.10, 95% CI [2.66, 3.05]), $t(118) = -1.07$, $p = .275$, 95% BCa CI [−0.41, 0.12], d = 0.20. Mean *d*' scores for positive words (M = 2.72, SE = 0.09, 95% CI [2.55, 2.90]) were only slightly, and not significantly lower than for the neutral words in the positive condition, $t(59) = -1.46$, $p = .145$, 95% BCa CI [−0.30, 0.05], d = 0.16. The mean *d*' for the negative words (M = 2.47, SE = 0.09, 95% CI [2.29, 2.64]) was significantly lower than for the neutral words in the negative condition, $t(59) = -5.38$, $p = .001$, 95% BCa CI [−0.73, −0.31], d = 0.80. Discrimination was not significantly different for negative compared to positive words, $t(117) = 1.80$, $p = .074$, 95% BCa CI [0.01, 0.50], d = 0.45.

There was no significant interaction between elaboration and word type on *d*' scores, $F(1, 110) = 0.47$, $p = .493$, $\eta_p^2 = .00$, nor was there a significant 3 way interaction between elaboration, word type and affect condition, $F(1, 110) = 0.12$, $p = .734$, $\eta_p^2 = .00$. The main effect of rotation on *d*' was not significant, $F(1, 110) = 0.62$, $p = .431$, $\eta_p^2 = .01$, nor did it interact significantly with affect condition, $F(1, 110) = 0.27$, $p = .604$, $\eta_p^2 = .00$, or elaboration, $F(1, 110) = 0.09$, $p = .760$, $\eta_p^2 = .00$. The three-way interactions between rotation × affect condition × elaboration, $F(1, 110) = 0.10$, $p = .753$, $\eta_p^2 = .00$, rotation × word type × elaboration, $F(1, 110) = 1.43$, $p = .235$, $\eta_p^2 = .01$, and rotation × word type × affect condition, $F(1, 110) = 0.66$, $p = .417$, $\eta_p^2 = .01$, were all non-significant, as was the four way interaction, $F(1, 110) = 0.40$, $p = .530$, $\eta_p^2 = .00$.

Response Bias: Criterion *C* analysis. *C* scores were not significantly different to 0 when lists were examined by type or valence condition (see Table 1). This indicates no significant bias in responding with ‘yes’ or ‘no’. However, significant conservative response bias was shown for neutral words in the positive, low elaboration condition, (*M* = 0.21, *SE* = 0.06), *t*(28) = 3.36, *p* = .002, 95% BCa CI [0.09, 0.32], *d* = 0.53, and negative words in the low elaboration condition (*M* = 0.15, *SE* = 0.06), *t*(29) = 2.65, *p* = .013, 95% BCa CI [0.04, 0.26], *d* = 0.70. In the high elaboration condition, significant liberal bias was shown for positive words, (*M* = −0.20, *SE* = 0.06), *t*(30) = −3.57, *p* = .001, 95% BCa = [−0.32, −0.09], *d* = 0.87.

Table 1.

Mean C scores by list type and valence condition with t-test showing whether C scores differ significantly from 0.

Word Type/ Condition	Criterion *C* *M* (*SE*)	95% BCa CI lower	95% BCa CI upper	*t*	*df*	*p*	*d*
Emotional	−0.03 (0.03)	−0.10	0.04	−0.96	118	.339	0.15
Positive	0.09 (0.05)	−0.18	0.01	−1.84	59	.070	0.43
Negative	0.02 (0.05)	−0.08	0.12	0.17	58	.614	0.09
Neutral	0.03 (0.03)	−0.03	0.08	0.83	118	.409	0.16
Positive	0.04 (0.04)	−0.05	0.13	1.01	59	.318	0.39
Negative	0.01 (0.04)	−0.07	0.09	−0.51	58	.900	0.09

Age and gender scores did not significantly predict response bias (*p*s = .891, and .479, respectively).[27] A 2 × 2 × 2 × 2 (Word type [emotive, neutral], Affect [positive, negative], Elaboration [low, high], Rotation [emotional first, neutral first]) mixed ANOVA with *Bonferroni* adjustments was conducted on *C* scores.

Response bias for emotive words (*M* = −0.03, *SE* = 0.03) was not significantly different than for neutral words (*M* = 0.03, *SE* = 0.03), *F*(1, 111) = 2.87, *p* = .093, η_p^2 = .03. There was no significant difference in *C* scores between positive (*M* = −0.02, *SE* = 0.04) and negative affect conditions (*M* = 0.02, *SE* = 0.04), *F*(1, 111) = 0.44, *p* = .508, η_p^2 = .00. However, word type and affect did interact significantly, *F*(1, 111) = 5.24, *p* = .024, η_p^2 = .05, with a slight liberal response bias (more likely to respond ‘yes’ than ‘no’) for positive words (*M* = −0.08, *SE* = 0.05), and a conservative response bias (more

27 Full covariate analysis is shown in supplementary materials.

likely to respond ‘no’ than ‘yes’) for neutral words in the positive affect condition (*M* = 0.05, *SE* = 0.04). A paired *t*-test showed that this difference was significant, *t*(59) = −3.26, *p* = .002, 95% BCa CI [−0.22, −0.05], *d* = 0.40. The difference between negative (*M* = 0.03, *SE* = 0.05) and neutral words (*M* = −0.01, *SE* = 0.04) in the negative affect condition was not significant, *t*(58) = 0.35, *p* = .730, 95% BCa CI [−0.09, 0.14], *d* = 0.11.

Elaboration instruction significantly affected response bias, *F*(1, 111) = 19.23, *p* = .000, η_p^2 = .15, with conservative bias observed in the low elaboration (*M* = 0.11, *SE* = 0.04) and liberal bias in the high elaboration condition (*M* = −0.11, *SE* = 0.04). Elaboration instruction did not significantly interact with word type, *F*(1, 111) = 0.49, *p* = .485, η_p^2 = .00, or affect condition, *F*(1, 111) = 0.98, *p* = .323, η_p^2 = .01. The elaboration × affect condition × word type interaction was also not significant, *F*(1, 111) = 2.64, *p* = .107, η_p^2 = .02.

Rotation order did not significantly affect response bias, *F*(1, 111) = 0.24, *p* = .623, η_p^2 = .00, and did not significantly interact with affect condition, *F*(1, 111) = 0.01, *p* = .905, η_p^2 = .00, or elaboration instruction, *F*(1, 111) = 0.35, *p* = .553, η_p^2 = .00. However, the interaction between rotation order and word type condition was significant, *F*(1, 111) = 5.52, *p* = .021, η_p^2 = .05. For those who received the emotional words first, there was no significant difference in bias between emotional words (*M* = 0.02, *SE* = 0.05) and neutral words (*M* = 0.00, *SE* = 0.04), *t*(57) = 0.47, *p* = .640, 95% BCa CI [−0.06, 0.11], *d* = 0.07. However, those who viewed the neutral words first showed a significant difference in bias between word types; with a tendency towards a liberal response bias for emotional words (*M* = −0.08, *SE* = 0.05) and towards a conservative response bias for neutral words (*M* = 0.06, *SE* = 0.04), *t*(60) = −2.88, *p* = .005, 95% BCa CI [−0.23, −0.05], *d* = 0.36.

The three-way interactions between rotation × affect condition × elaboration, $F(1, 111) = 3.63$, $p = .059$, $\eta_p^2 = .03$, rotation × word type × elaboration, $F(1, 111) = 1.32$, $p = .254$, $\eta_p^2 = .01$, and rotation × word type × affect condition, $F(1, 111) = 0.22$, $p = .640$, $\eta_p^2 = .00$, were non-significant, as was the four way interaction, $F(1, 111) = 0.12$, $p = .725$, $\eta_p^2 = .00$.

Discussion

This experiment aimed to explore mechanisms underlying observed negativity effects identified in Experiments 2 and 3 (Chapters III and IV) by manipulating elaboration levels. It was hypothesized that: (1) as levels of elaboration increase, the valence effect will diminish; (2) discrimination will increase with age; and (3) there will be no gender effects.

Key findings showed: (1) a significant effect of valence on levels of discrimination; (2) a significant effect of valence on levels of response bias; (3) a significant interaction effect of word type and rotation on response bias; (4) a significant effect of semantic elaboration on response bias; and (5) a significant relationship between age and discrimination. The effects of valence on memory will first be considered, followed by a discussion of the effects of manipulating levels of semantic elaboration on emotional memory. A brief discussion of moderators follows, with considerations of implications and suggestions for future exploration.

Valence Effects on Recognition

Valence was successfully manipulated between the DRM lists. The lists have previously been validated as positive, neutral and negative in valence (Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008; Budson et al., 2006). Valence scores taken from the Warriner, Kuperman, and Brysbaert norms for 13,915 English lemmas (2012), and children’s word rating scores collected within the high elaboration condition show that the emotionally positive, neutral, and negative word lists are significantly different from each other in the expected directions.

A significant, medium sized effect of valence on discrimination was observed as a preference away from negative words; evidenced by the interaction between word type and valence condition. This supports findings from previous studies using similar DRM paradigms with children (Howe, 2007; Howe et al., 2010, Experiment 2, 4, and 5; Moradi et al., 2000; Neshat-Doost, et al., 1998). No significant preference toward or away from positive words was shown, although there was a tendency away from positive words, in line with Moradi et al., (2000) and Nehsat-Doost et al. (1998).

There was also a difference in response bias between negative and neutral words. Where neutral words were presented first there was a slight conservative bias shown towards the neutral words, but a slightly higher liberal bias shown towards the emotional words. This liberal response bias would have resulted in a higher amount of false alarms for the emotional that neutral words, in line with previous findings (Budson et al., 2006; Howe et al., 2010). There was no significant difference in response bias between the neutral and emotional words when emotional words were presented first. Together with discrimination results, this shows that children are both more inclined to respond positively (say they have seen emotional words before), in particular for positive words, or when they follow neutral words, and this is a separate effect than the lower accuracy seen for negative compared to neutral words. It would thus appear that negative stimuli are processed less effectively within memory, whereas children are more likely to report positive memories, regardless of whether they are accurate.

There was a small, but non-significant effect of valence on discrimination between positive and negative words, with positive discriminated more accurately than negative. This is in line with findings of a significant effect reported by Brainerd et al

(1998), Moradi et al. (2000), and for visual stimuli when presented with a congruent story in Experiment 2 (Chapter III). There was no valence effect on response bias in this cell of comparison.

Children exhibited slight liberal bias towards positive items, and conservative bias towards neutral items. Although the difference between positive and neutral words was significant, significant levels of response bias were not observed in either condition independently. Additionally, those who viewed the neutral words first showed a significant difference in bias between word types; with a tendency towards a liberal response bias for emotional words and towards a conservative response bias for neutral words.

The difference in findings between Experiment 3 and 4 could be due to several reasons. Firstly, the near ceiling effect appeared even greater in Experiment 4 than in Experiment 3, despite a distractor task being introduced in an attempt to increase task difficulty. This effect held across the positive and neutral conditions. This could be due to children completing the recognition tasks for fewer lists, and so becoming less fatigued or bored. Alternatively, it could be due to the different procedure; there was a central fixation point displayed between each word, meaning words were presented at a slower rate, and words were displayed for 2 s in the initial presentation phase before children could click to rate the words on ease of spelling or emotion evoked. Words continued to be displayed until children had clicked their spelling of emotion rating after this initial 2 s, meaning they were displayed for longer, overall in Experiment 4, than Experiment 3. Additional rehearsal could have been taking place during this extra time. Response speeds were not recorded, but it was evident during sessions that children routinely attempted to respond within the first 2 s of words being displayed, and therefore had to continue looking at the screen and wait until they were allowed to respond in accordance with the programming design. As discussed in Chapter IV, the paradigm in Experiment 3 also included presentation of a story prior to viewing words. This could have led to a difference in processing as children sought to match up the words with contextual information from the stories in Experiment 3, but not in the current experiment (Haviland & Clark, 1974). Lastly, Experiment 3 was likely much more underpowered than Experiment 4.

The limitations with the word lists discussed in Experiment 3 also apply when interpreting the findings within this chapter; there are several significant differences between the word lists, other than valence, which present as confounding variables; word lists differed significantly in arousal, age of acquisition, and imageability. All of these are characteristics are likely to have affected discrimination. Although arousal positively predicts memory performance (Lang et al., 1995), this was not observed in the current study with the least arousing neutral words discriminated most accurately, followed by positive words, with the most arousing negative words discriminated least accurately. Age of acquisition also positively predicts recognition memory (Cortese, McCarty, & Schock et al, 2014). Findings of this experiment do not match existing literature, with negative words (acquired the latest) discriminated the least accurately.

Imageability positively predicts recognition memory (Cortese et al, 2014). Negative words scored significantly lower than neutral words on imageability, but no different to positive words, meaning the neutral words should have been discriminated more accurately than the negative words. This pattern was observed, with significantly higher discrimination observed for neutral than negative words. A trend for higher discrimination for neutral, compared with positive words was also observed, in line with a trend for lower imageability scores for positive, compared with neutral words. This means that imageability could be driving the findings of the current experiment, in combination with, or instead of valence (as opposed to arousal and age of acquisition in Experiment 3). These particular word lists were chosen to extend on Experiment 3, but in future studies, word lists should be controlled on all characteristics except valence, to truly tease these findings out. Further research should be carried out to establish why arousal and age of acquisition appeared to drive discrimination effects in Experiment 3, but imageability in Experiment 4.

Effects of Semantic Elaboration on Emotional Memory

It was hypothesized that as explicit elaboration increased, valence effects would decrease (hypothesis 1). Semantic elaboration levels did not significantly affect discrimination rates, contrary to the hypothesis and findings from a previous DRM experiment, where more correct items were recalled in an immediate free recall task when word lists were encoded semantically compared to non-semantically (Toglia, Neuschatz, & Goodwin, 1999). However, children did exhibit significantly more liberal response bias in the high elaboration condition, and conservative bias in the low elaboration condition. This effect was driven by significantly conservative bias for neutral words in the low elaboration, but significantly liberal bias for the negative words in the high elaboration condition. Where response bias is observed, true sensitivity is underestimated (Georgeson, n.d.). Taken with low power (around 15 observations per cell), it is likely that any true effects of level of elaboration were obscured in the current experiment.

Differences in findings between the current experiment and findings of Toglia et al. (1999) could be explained by experimental design differences. Firstly, the current experiment included children, whereas Toglia et al. (1999) used an adult sample. Research has shown a cognitive shift in false memory occurs somewhere between 7-11 years of age (Brainerd et al., 2010). Secondly, Toglia et al (1999) employed a free recall design, where participants were instructed not to guess, whereas the current study utilised a recognition task. Differences between rates of false alarms in recognition tasks and false recall have previously been reported (Howe et al., 2010). Recall tasks activate four separate brain areas more fully than recognition tasks (Cabeza et al., 1997), possibly indicating that these additional steps of accessing information could be where semantic elaboration has an effect on false memories.

In the low elaboration condition, children were instructed to rate words based on the difficulty of spelling each word. This required children to access lexical and phonological information about each word, which would activate only other words with similar spellings or sounds. This is equivalent to a shallow level of encoding. In the high elaboration condition, children were instructed to rate how the word made them feel. This required children to think about the meaning of the word, and in what situation or context they would encounter the word. This process would have activated a much wider network of other related words and concepts in their memory, and may have increased feelings of recognition for words in the recognition task that had not been presented in the learning phase, encouraging a '*yes*' response more frequently (Demb et al., 1995), accounting for the increase in liberal bias.

Moderators of Emotional Memory

As expected, rates of discrimination increased with age (hypothesis 2). Working memory becomes more efficient through childhood, resulting in a greater working memory span and the ability to successfully encode and remember more items (Gathercole, 2004). These age effects are in line with findings of other similar DRM studies (Brainerd, 2010; Howe, 2007; Howe et al, 2009, Howe et al; 2010; Quas, 2013).

There were no interactions between age and valence effects on discrimination or response bias, as in Experiment 3 of this thesis. However, in Experiment 2 of this thesis, as well as in existing literature (Brainerd et al., 2010), recognition increased with age for the emotionally positive material at a faster rate than for the emotionally negative material. Additionally, recognition for negative pictures was found to increase as age increases, but remain stable for neutral images (Cordon et al., 2013). The mixed pattern of results provides inconclusive evidence as to the precise relationship between age and valence effects, and under what circumstances this effect is borne out.

In line with hypothesis 3, gender was not significantly associated with discrimination or response bias, as in Experiments 2 and 3, and in line with previous research using word list (e.g. Howe, 2007; Howe et al, 2009; Howe et al., 2010; Quas, 2015) and images (Cordon, 2013). Although a gender effect was identified in Experiment 1 of this thesis, this is thought a spurious result; no explanation with a firm theoretical underpinning could be identified.

Conclusions

The present experiment provides support for existing findings of a preference away from negative stimuli in discrimination, when presented as word lists in a DRM paradigm. A tendency for lower discrimination for positive than neutral, and for negative than positive words was also found, in line with previous findings. Discrimination increased with age, but no differences were found between gender (hypotheses 2, and 3). Neither interacted with valence effects. The processing of emotional words is thought to result in increased activation of semantic networks in memory (Demb et al., 1995), triggering an increase in false memories for related words or words of a similar valence. It was hypothesized that increasing explicit semantic elaboration would lower valence effects by ‘evening the playing field’, as it were (hypothesis 1). Although no effects of elaborative instruction were seen on discrimination, deeper encoding led to a higher liberal response bias, with bias differing by valence, depending on elaboration condition. This interaction can be explained by a positive-negative asymmetrical bias in information processing (Lewicka et al., 1992).

In future experiments, longer word lists should be utilised, with a faster presentation speed, and no central fixation point displayed between each item. These changes would reduce the amount of time available for active rehearsal, to reduce near-ceiling effects. Word lists should also be matched on other aspects such as arousal, imageability and age of acquisition. Increasing power and implementing a much longer distraction task would also make it more likely that valence effects would reliably be found in the control conditions; given a minimum delay of 20 minutes is typically needed to produce emotional enhancement effects in memory (Kleinsmith and Kaplan, 1963). Including lists of related and unrelated emotional and neutral words would also allow differential effects of cohesiveness and semantic relatedness to be teased apart more fully.

Chapter VI: Discussion

Aims and Methodologies

The aims of this thesis were: (1) to explore whether environmental influences (specifically “scary” TV) are a plausible explanation for increasing levels of anxiety in typically developing children; (2) to explore the relationship between trait anxiety and emotional preferences in memory; and (3) to explore whether typically developing children show a preference for emotionally valenced stimuli in long term memory. A secondary line of enquiry was to explore age and gender as potential moderators of emotional memory biases, and to investigate potential interaction effects of stimulus types and context of presentation.

This conclusion chapter first provides a brief recap of the design of the meta-analysis and four experiments in this thesis. It proceeds to first discuss the internalizing effects of scary TV from Chapter II (research aim 1) and how these might relate to emotional memory preferences, followed by a discussion of the effects of emotionally valenced stimuli identified within the experimental chapters (research aim 3). Interactions between context and stimulus types are considered alongside these findings, followed by a discussion of depth of encoding as a potential mechanism for valence effects on memory as secondary lines of enquiry. Age, gender, and trait anxiety (research aim 2) are presented separately because these were not manipulated as part of the experimental design. Finally, contributions and limitations of the experiments, and thesis as a whole are discussed, along with future directions for research.

Meta-analysis Design

The first aim was addressed by conducting a meta-analysis of the effects of scary TV on internalizing emotions, including fear/ anxiety, sadness/ depression and PTSD amongst others, in children (Chapter II). Experimental and correlational studies were included; experimental studies were included only when the study design incorporated a control group of no TV viewing, or baseline measures. This allowed a weighted, standardized effect size to be calculated, to complete a systematic review of existing literature exploring these effects. In addition, ten potential moderators were entered into the analysis as a multi-level nested model, based on theoretical hypotheses and previous experimental findings. These were grouped into *study characteristics*, *individual characteristics*, and *media characteristics*.

Experimental Designs

Experiments 1-3 addressed the second research aim by measuring levels of trait anxiety (one of the internalizing outcomes in the meta-analysis), and exploring whether the relationship between anxiety and memory performance differed between materials of positive, neutral or negative valence. Trait anxiety was not measured in Experiment 4 because no reliable effects had been borne out in the previous experiments. All experiments explored research aim 3, although only a comparison between negatively and positively valenced materials could be made in Experiment 1, and the recognition task in Experiment 2. The recall task in Experiment 2, and Experiments 3 and 4 also included comparisons between valenced and neutral stimuli to explore negativity and positivity preferences.

Experiment 1 (Chapter III) explored the effects of valenced TV clips on memory performance. Interaction effects of stimulus type were also explored. An emotionally positive and an emotionally negative clip were shown to typically developing primary school aged children, followed by a free and prompted recall task, and a recognition task. However, the main manipulation of valence was unsuccessful, in addition to the inclusion of only one exemplar of each valence, and no neutral comparison condition. Conclusions based on findings from Experiment 1 were therefore limited.

Experiment 2 (Chapter III) built on the design and trends identified in Experiment 1 by presenting still images from each clip, paired with either a positively or negatively valenced version of the narratives adapted from the same TV clips. Children were assigned to either a congruent condition, where the images from each clip were presented with a narrative that matched the valence of the images, or an incongruent condition. This assignment allowed the effects of stimulus modality, and congruency/ context on differently valenced materials to be tested. It also permitted for a portion of the positively and negatively valenced version of each narrative to consist of shared, neutral text. This allowed negativity and positivity preferences in recall memory to be explored (Research aim 3).

Experiment 3 (Chapter IV) was designed to further explore the context effects seen in Experiment 2. Children were shown a valenced or neutral story, followed by emotionally negative, neutral and positive DRM word lists. The story valence provided emotional context, to compare to performance on recognition for word lists of the same or different valence.

Experiment 4 (Chapter V) was designed to test the role of elaboration in memory preferences, as a possible mechanism for the effects apparent in Experiments 1-3. The word lists from Experiment 3 were used for continuity, and level of elaboration was manipulated using ratings tasks. Children rated words on ease of spelling in the low elaboration condition, and depth of positive or negative feeling evoked in the high elaboration condition.

Internalizing Effects of “Scary” TV

Chapter II presented a meta-analysis examining the internalizing effects of scary TV on children. The meta-analysis found a small but consistent effect of scary TV on children’s internalizing responses. This finding is in line with Twenge (2000), who hypothesized that an increase in TV viewing could be leading to an increased perception of overall threat in the world, due to the amount of threat content regularly broadcast. The implications of these findings may apply to policy makers, clinicians and parents; ratings guidelines should be updated to consider how much ‘scary’ content is included in TV aimed at both children and adults, in addition to violent content. Parents may also wish to consider more carefully, what content they consider suitable for their child. Children, particularly those aged under 10 years, should be monitored for signs of internalizing emotion (for example, fear/ anxiety, sadness, sleep disturbances) during and after viewing. Content causing symptoms such as these may not be suitable, and it is not recommended that parents expose children to any content they find highly distressing. Although systematic desensitization or graded exposure therapy can be effective forms of treatment to reduce anxiety caused by a particular stimulus or environment (Powers & Emmelkamp, 2008), therapies should only be administered by qualified professionals.

Anxiety has been linked to distortions in attention, interpretation and memory (Muris & Field, 2008). Whilst attention and interpretation effects have been fairly well established, the link between memory and anxiety is more tenuous. The effect of valence on memory was therefore explored in the experimental chapters, as a potential cognitive mechanism leading to greater fear for negative stimuli.

Moderator analysis in the meta-analysis revealed that experimental designs resulted in a significantly larger effect size than correlational studies. Experimental paradigms were therefore employed in Chapters III-V, as this would enable the mechanisms behind the effect to be studied more easily. A significantly larger effect size was also reported when all children in the study sample were less than 10 years of age. For this reason, Experiments 1-4 included primary school aged children (7-11 year olds). Existing literature has shown that children of this age process information more visually, and are more easily scared by mythical or magical content (Harrison & Cantor, 1999). For this reason, TV clips of a programme based in a fictional, magical land were chosen for Experiments 1 and 2, with memory scores split into visual and narrative categories.

Valence Effects on Memory

Positivity and Negativity Preferences

No preferences toward positive or negative, compared to neutral stimuli were observed in any of the experiments, in recognition or recall tasks. This was somewhat surprising where pictorial or narrative stimuli were used, given previous research (Cordon et al., 2013; Van Bergen et al., 2015).

A preference away from emotional (toward neutral) stimuli was found in the recall task where the two sources of information were incongruent (Experiment 2), and a preference away from negative words was found in the DRM paradigm that included the use of an encoding task (Experiment 4). These were medium sized effects, with no response bias evident. Valence was successfully manipulated in Experiment 4 and the manipulation was partially successful in Experiment 2. Taken with the effect size, significance of findings, and lack of response bias, this indicates they are likely to be genuine effects.

However, this valence effect was not consistent across all four experiments, so cannot be considered a general effect of valenced material. There is no obvious similarity between the designs of these two experiments. It is unlikely due to an extraneous variable relating to the properties of the stimuli (such as level of arousal), because the same effect was not found in Experiments 3, which used the same word lists as Experiment 4, whilst the stimuli in Experiment 2 and 4 were markedly different from each other. Nonetheless, there were limitations with both stimulus sets in regards to extraneous variables.

The image sets in Experiment 2 were taken from the clips in Experiment 1. These were from the same TV show, and were matched on several other key variables such as arousal,[28] main characters, and length, for example. There may have also been other key differences between stimuli sets in Experiment 2 not considered during design such as novelty and cohesiveness, which can influence memory performance (Keenan, Baillet, & Brown, 1984; Kishiyama & Yonelinas, 2003). Crucially, valence ratings did not differ significantly between the two versions of each story, or between each set of images, resulting in only a partially successful manipulation of valence. However, they differed significantly for the ‘druid’ versus ‘monster’ condition overall, in the direction expected according to the valence of images, which remained constant regardless of congruency condition. It is possible that presenting two sets of stimuli simultaneously affected their ratings due to contextual cues. For example, children reported feeling significantly more positive towards neutral words when they were presented alongside positive words, than when identical neutral words were presented alongside negative words in Experiment 4. The absence of the expected interaction effects of story condition, modality, and valence is not considered to negate preference away from emotional material outlined above, however. Two other consistent patterns of findings were observed, showing valence effects between positive and negative stimuli, and an

---

[28] See Appendix for analysis of excitement ratings for Chapter III, Experiments 1 and 2.

interaction effect between modality and congruency on memory. The experiment may have lacked enough power to identify any true interaction effects in valence ratings.

As discussed in Chapters IV and V, word list stimuli in Experiment 3 and 4 were more highly controlled on valence and other word characteristics such as length and frequency, but the lists did vary on some key characteristics; arousal, age of acquisition, and imageability. All three of these characteristics are related to performance on recognition tasks, and therefore could be partially driving the effects which were found (Cortese et al., 2010; Lang et al., 1995). The pattern of results does not match that expected if arousal or age of acquisition were driving the effect. And, as already highlighted, results differed between Experiments 3 and 4, meaning these other list traits may take on varying roles depending on other experimental factors.

These significant findings on discrimination correspond to the majority of previously published research (Howe, 2007; Howe et al., 2010; Moradi et al., 2000; Neshat-Doost et al., 1998). Howe (2007) suggested a potential cognitive mechanism which children may employ, resulting in this valence effect in recall; children might use an 'editing function' for negative but not neutral information whereby they can filter out both true and false negative items. This is consistent with literature reporting that children are often reluctant to talk about negative events, and are less likely to assent that negative false events occurred than to assent that positive or neutral false events occurred (Bruck & Ceci, 1999).

Experiment 4 incorporated a recognition rather than recall task, limiting the explanatory power of Howe's theory (2007). Furthermore, negative material tends to be more semantically dense than neutral material (Talmi & Moscovitch, 2004). If negative emotional items are more interrelated than neutral items, they may be processed with greater depth or richness (e.g., Dewhurst & Parry, 2000; Ochsner, 2000), increasing relational processing. This would lead to an increase in discriminability as semantic relatedness increases (Dewhurst & Parry, 2000; Ochsner, 2000; Talmi, Luk, et al., 2007a; Talmi & Moscovitch, 2004; Talmi, Schimmack, et al., 2007). It is unclear why this was not found in the current experiments, but could potentially be due to other factors in the experiment such as alternative characteristics of the word lists which can impact memory performance, and differences in study design and methodology.

Response bias. There were also some effects of valence on response bias, independent of the valence effects outlined above. Children exhibited a higher propensity to respond with 'yes', or lower propensity to respond with 'no' for emotionally valenced, compared to neutral words in both experiments that included a neutral group (3 and 4). This effect was not observed for both positive and negative words in every cell of the highest order interactions, but was fairly consistent. The effect was small for positive, but large for negative words. This means children are using a criterion other than genuine discrimination to make decisions. An increased bias towards responding ‘yes’ can result in higher false alarms, which may have driven any subsequent discrimination effects. False alarms are not analysed separately throughout the thesis, but it is possible that an increase in false alarms can lower overall discrimination enough that differences between conditions are no longer significant, even if one condition has a significantly greater hit rate. This is an alternative explanation for the lack of observable negativity and positivity preferences within discrimination.

Positive vs. Negative Stimuli

Overall, memory performance for negative stimuli exceeded that for positive, across the experiments presented within this thesis. This effect was evident only for narrative or verbal items, but not visual items, indicating that modality interacts with valence effects on memory. This was reliably demonstrated in Experiment 2 across recognition and recall, for both the congruent and incongruent narrative/ image set conditions. However, the effect size was much larger when the narrative/ image sets were congruent. A small effect of valence was also found for discrimination of words in Experiment 3. An interaction with rotation qualified this main effect, whereby this was only significant when negative lists were presented first. Experiments 1 and 2 did not uncover the same effect between positive and negative materials. However, the manipulation of valence in Experiment 1 was unsuccessful, meaning valence differences would not be expected. There is no obvious reason why a trend in the opposite direction was found in Experiment 4, other than the addition of an encoding task.

Notably, very few differences in response bias were identified between positive and negative conditions across the experiments. Children displayed less conservative bias for negative than positive words in Experiment 3, but only in one rotation condition. Although this mirrors the valence and rotation effect on discrimination, it was only found in a handful of children, and cannot be generalized. Decision criterion does not appear to differ substantially for negative, compared to positive materials.

Greater memory performance for negative than positive stories has also been identified in previous literature (Baltazar et al., 2012; Miller & Sperry, 1988; Van Bergen et al., 2015). One explanation for these findings, presented by Van Bergen et al (2015) is that negative emotions typically involve greater complication or goal failure than positive emotions (O'Kearny & Dadds, 2005). Negative experiences usually provoke increased effort towards understanding, and more effortful cognitive processing (Habermas, Meier, & Mukhtar, 2009; Lagattuta, 2014). This increase in cognitive processing would subsequently increase memorability of the negatively valenced stories

(Lagattuta, 2014). Although Experiment 3 used word stimuli rather than stories, stories were presented immediately prior to the negative word lists, when this valence effect was observed. Experimental evidence shows that when new stimuli are presented, individuals first search their memory for antecedent information that matches the new stimuli (Haviland & Clark, 1974). In the current experiment, children could have potentially searched the antecedent stimuli in memory (the story) for the negative stimuli, when they were presented first; resulting in a similar effect.

By comparison, there was only one condition where memory performance for positive material significantly exceeded that for negative; for visual stimuli in Experiment 2, when the narrative/image sets were congruent. This corresponds with a lower level of response bias for the positive, than negative visual items, regardless of congruency. However, this modality effect cannot be interpreted adequately in isolation. Negatively valenced narrative was also discriminated at a significantly greater rate than the images, with no such difference observed for the positively valenced material.

One possible explanation is the effect of valence on attention. Attention is considered a limited resource, and material needs to be selected to be attended to; various models provide differing hypotheses of how selection is undertaken. For example, Broadbent hypothesized that a specific physical characteristic of one channel of information is selected to attend to, without processing semantic content (Broadbent, 1958). Deutsch and Deutsch (1963) extend upon Broadbent’s model and propose a shifting reference standard, whereby the filter takes up the level of the most important arriving signal. This means that when one signal is perceived as more important, it will be preferentially attended to. In the case of the negative condition, stress can trigger coping behaviours such as information seeking (Hoffner, 1993; Miller, 1987). If the negative narrative provides a richer source of information, attention may be directed towards that, at the expense of the images. Increased attention can lead to an increase in memory performance through various mechanisms (for a review see Kensinger, 2004). For positively valenced material, this attention split may be more even between the two sources, resulting in no significant difference in discrimination. Therefore, when comparing discrimination for positive and negative visual items, discrimination appears greater for positive visual items. Discrimination for the images could be tested separately to test this hypothesis, as well as measuring attention during the discrimination task.

Depth of Encoding

It was originally anticipated that introducing a deep encoding condition in Experiment 4 would both enhance memory performance (Toglia et al., 1999), and potentially reduce valence effects (Talmi & Moscovitch, 2004). Neither of these effects were borne out, but a difference was apparent in response bias; liberal response bias was evident in the high elaboration condition, whilst conservative response bias (of a lower value) was evident in the low elaboration condition. This effect was driven by significantly conservative bias for neutral words in the low elaboration, but significantly liberal bias for the negative words in the high elaboration condition. It is hypothesized that in the higher elaborative task, a denser network is activated. This would lead to lower discriminability in a subsequent recognition task (Demb et al., 1995). This initially seems to discredit one of the possible explanations offered in response to effect of greater recognition for semantically denser material. However, a crucial difference here is that Experiment 3 involved a story context, providing a schema to limit the spread of activation to unrelated items. This could therefore be an extra step, modifying valence affects. Alternatively, findings from either experiment may have been spurious rendering one explanation null. Further research would need to be conducted, potentially with more divergent levels of encoding tasks to test these theories more fully. For example, the shallow encoding task could focus on visual aspects of the word displayed, such as colour of font.

Contribution of Other Factors on Memory Performance

Modality

The interaction effects between valence and modality have been discussed above. However, modality also appears to interact with congruency independently of valence. In Experiment 3, the narrative items were discriminated significantly better than the visual items for both story conditions. Both were large effects. This suggests that when an element of uncertainty or stress in introduced, the richest source of information is sought out and attended to, regardless of its valence; in the same way as in the congruent negative condition. This could then lead to an increase in memory (Kensinger, 2004).

Context

In addition to the congruency effects discussed above, mood was manipulated in Experiment 3 prior to presenting the word lists. Lists of a congruent valence would be remembered better, with a potential interaction between this effect and valence. No context effects were found, despite successful mood manipulation. This is most likely due to ceiling effects, low power of the study, or only partial success in manipulating mood.

Moderators

Three potential moderators were measured and included in analytical models across the four experiments: age, gender and trait anxiety. Overall, there was a trend for memory performance to increase with age across the experiments, but no reliable effects of gender or trait anxiety on memory performance were found. None of the three variables appeared to moderate emotional memory reliably; a tentative finding is that age may moderate recognition and recall memory performance for emotionally positive and negative materials differentially depending on whether the stimuli is semantically dense.

Age

In three of the four experiments (Experiments 1, 2 and 4) discrimination increased significantly with age. This held across all stimulus types; both visual and verbal aspects of TV clips, static images, aurally presented narratives and word lists. The proportion of narrative material correctly recalled in the monster condition of Experiments 1 also increased significantly with age, as did the number of false narrative details recalled overall in Experiment 1.

A general increase in memory performance was expected with age (Gathercole, 1998) although findings have been mixed in previous literature. Howe (2007) and Howe et al. (2010) analyse recognition and recall for true (target items) and false items (critical lures) separately, reporting that both increase with age. In contrast, Cordon et al (2013) report an increase in discrimination with age, whilst others report no significant changes with age (Moradi et al., 2000; Neshat-Doost et al., 1998; Van Bergen et al., 2015). An increase in both true and false information recognized or recalled could lead to an increase in discrimination, or no difference, depending on how strong each relationship was. The current findings suggest that memory performance for correct items increases at a faster rate than that for false items.

There was an interaction between age and valence on discrimination in Experiment 2, whereby discrimination increased with age for positive, but not negative material. This may be because negative false alarms were increasing, along with hit rates, but positive false alarms did not. This would need to be tested with further analysis. This effect was only found once, and cannot be considered a general effect across memory. The implications of this are that valence effects were generally stable across the age ranges included in the four experiments. Valence effects would therefore have either developed before age 7 years, or remained constant from birth.

Despite existing literature showing a negativity preference in attention from a very young age, and therefore providing evidence against acquisition models (Field & Lester, 2010) these studies primarily used evolutionarily threatening stimuli such as snakes and spiders. The stimuli used throughout this thesis include mythical creatures (Experiments 1 and 2) and words lists not related to survival or mortality such as *cry*, *sweet* and *music* (Experiments 3 and 4). It is possible that memory preferences for some types of emotional stimuli are acquired and some are present from birth. Alternatively, it may be the case that other factors combine to moderate these preferences. Experiments of a similar design would need to be repeated with younger age groups to explore whether memory preferences for this type of stimuli were explained best by integral, acquisition or moderation models of fear acquisition.

There were no significant effects of age on response bias in any of the experiments, or number of false memories recalled in Experiment 1 and 2. Most DRM studies with children have not reported response bias, but Cordon et al. (2013) found children exhibit more liberal response bias than adults. Cordon et al. used purely pictorial stimuli, so it is not possible to directly compare with the experiments in this thesis.

Gender

In Experiment 1 boys could discriminate signals from noise to a significantly greater extent than girls. However, this effect was not replicated through Experiments 2-4. There was no significant effect of gender on response bias through any of the experiments, and no significant effect of gender on levels of true, false or visual recall in Experiments 1 and 2. Most child DRM studies either do not explore gender (Brainerd, Holliday, & Reyna, 2004; Brainerd, Reyna, & Forrest, 2002; Howe et al., 2010; Howe & Derbish, 2010b; Metzger et al., 2008), or report no significant gender effect (Cordon et al., 2013; Howe, 2007). As discussed in Chapter III, this anomaly could be due to a particular feature of the stimulus. Streicher and Bonney (1974) note that the fantasy genre is typically more appealing to boys, so they are more likely to watch it. In addition, familiarity with fantasy characters can increase recognition of new characters from the same series in visual short-term memory (Xie & Zhang, 2016). However, there were no significant gender differences reported in how much boys and girls liked watching *Merlin*, levels of excitement after viewing the clips, or frequency of viewing *Merlin* at home, negating these potential explanations.

There is some limited evidence that females produce higher rates of false memories for negative stimuli than males (Dewhurst et al., 2012). Higher false alarm rates for females would mean lower discrimination compared to males, if hit rates were equal, mirroring the result shown in Experiment 1. However, Dewhurst et al. included an adult rather than child sample, and negative and neutral, rather than negative and positive stimuli. The authors suggest that women reflect on associations within negative lists to a greater degree than men and are more likely to generate the negative critical lures. If this held only for negative, rather than all emotional content, then an interaction would be expected between gender and valence, which was not found. Additionally, it is unclear why these results were found for just one set of results in the thesis, and only within one existing publication. The observed gender effect was therefore treated as a spurious finding.

Trait anxiety

Trait anxiety was included as a third potential moderator for Experiments 1-3. If memory preference for negative material (such as scary TV) is the mechanism for internalizing emotions, including anxiety, then trait anxiety should co-vary differentially for stimuli of negative, compared to positive or neutral valence.

In Experiment 1 there was an increase in the proportion of true memories recalled that matched the overall valence of the clip, as trait anxiety increased. No other effects of anxiety were found throughout Experiments 1-3. This suggests that more dominant information types are focused on as anxiety increases; but does not suggest any effects of valence. It may be that the relationship between trait anxiety and valence effects on memory only becomes large enough to identify an effect when trait anxiety reaches higher than average levels. For example, attention preferences are significantly higher in clinically anxious compared to non-anxious children (Dudeney et al., 2015). Alternatively, memory preferences towards negative content may not be the mechanism mediating the relationship between scary TV and anxiety.

Contributions and Limitations

Child Anxiety

This thesis demonstrates a reliable effect of scary TV on children's internalizing emotions. Whilst many researchers have warned of the dangers of TV viewing, particularly for children, much of the advice to date appears to have been based on qualitative evidence and anecdotal accounts of negative effects. Despite showing that negative symptoms increase with scary TV viewing, this finding should also be kept in context. The effect size was small, suggesting that for most children the viewing of scary TV is unlikely to cause substantial distress past that experienced in everyday life. However, moderator analyses offer some further insight; children under 10 years of age may be particularly susceptible to negative emotional effects, and are more likely to show physical symptoms of internalizing emotions. Policy makers and parents should bear this in mind when setting guidelines and assigning ratings to TV shows, and when deciding what appropriate viewing for their child is.

An increase in memory performance for negative, over positive narrative material partially supports the hypothesis that fear responses towards scary TV content could be moderated by valence effects in memory. However, an increase in memory for negative over neutral material would also be expected. By its nature, the meta-analysis included a mix of studies, including experimental and correlational studies, with both shorter and longer-term outcomes on internalizing symptoms. Dominant theories of emotional memory draw a distinction between emotional enhancement during short term and long term memory. A neuroimaging and neuropsychological review by Hamann (2001) concludes that emotional arousal influences memory via factors that act during memory encoding (attention and elaboration) and factors that modulate memory consolidation. The Memory Modulation Model proposes that stress hormones released during emotionally arousing situations modulate memory processes. McGaugh and Cahill (1996) and McGaugh (2004) provide literature reviews, concluding that the amygdala mediates the memory-modulating effects of these stress hormones and several classes of neurotransmitters when modulating the consolidation of long-term memories. The preferences identified within short term memory tasks have limited applicability to long term internalizing outcomes (such as the effects of PTSD two years after a bombing; Pfefferbaum, 2000). Time between exposure and internalizing response should be included as a moderator for future meta-analyses, and future experiments should seek to explore long term memory mechanisms to explore long term internalizing responses observed after TV viewing.

Children's Emotional Memory

The remainder of this section highlights some of the difficulties and limitations which were faced when exploring emotional memory throughout this thesis, and how these have in part been successfully navigated to add to the small, but growing body of literature focusing on children's emotional memory.

Difficulties in designing appropriate stimuli. Success of the main manipulation of valence varied between experiments. Experiment 1 and 2 sought to use more naturalistic stimuli than the majority of previously reported emotional memory studies with children, which have predominantly used picture (Cordon et al., 2013) or word lists (e.g. Howe, 2007; Howe et al., 2010; Moradi et al., 2000; Neshat-Doost et al., 1998; Quas et al., 1999). In psychological experiments there is always a trade-off between having tight control of conditions and exploring behaviour or cognition in the manner in which it would naturally occur in day-to-day life. The clips used in Experiment 1 and images in Experiment 2 were edited from TV episodes of *Merlin* (Hawes et al., 2009). Due to the fantasy and adventure element to the programme, it was not possible to edit an episode down into an entirely positive clip or negative clip, and a small proportion of both the positive and negative clips contained material of the opposite valence. Consequently, the levels of internalizing emotions reported after viewing each clip did not significantly differ. It was also difficult to control for other properties of the clip, such as familiarity, novelty, and arousal, although analysis suggested that children were equally familiar with both episodes of *Merlin* used in Experiment 1 and 2, and stimulus did not differ in arousal. Valence was successfully manipulated for the overall positive and negative conditions in Experiment 2, allowing for some consistent, if tentative findings on memory performance for naturalistic, but controlled narrative material to come to light.

Emotionally neutral material. The difficulty in generating appropriate neutral stimuli has been noted by previous researchers (Kensinger, 2009; Van Bergen et al., 2015). This was addressed somewhat in Experiment 2 by adding a neutral portion of narrative for each story, which remained constant between the negative and positive versions. This addition allowed for the naturalistic element of the stimuli to be retained, because the stories were matched to the images directly taken from the clips in Experiment 1, but also allowed for some comparison between memory performance for the emotional and neutral portions in each valence condition. However, this still did not provide an overall control condition, and for this reason, entirely neutral words lists were included in Experiment 3 and 4 as control conditions, at the loss of some ecological validity.

Other properties of word lists. Many other properties of words can affect memory performance in addition to valence (see Appendix, Chapters IV, V – Stimuli for a review). The word lists employed as stimuli in Experiments 3 and 4 differed in valence in the anticipated directions, but also differed on some other key aspects;

arousal, age of acquisition, and imageability. Lists should be carefully constructed in future experiments, to ensure that these characteristics do not vary between conditions.

Limited items within trials. An additional potential limitation is the use of only one positive and one negative clip (Experiment 1), and likewise one positive and one negative set of images (Experiment 2). Although the clips contained multiple scenes, and the image sets contained multiple still shots, they represented only one overall story. It may be suggested that any differences in memory performance between the two clips in Experiment 1 could be due to a confounding variable of the specific clip chosen. The TV clips were complex in nature, including multiple characters, actions and events, and themes in each clip. For example, the negative clip includes two negative characters (a witch and a monster), various scenes with negative actions including a monster hatching, examining a dead body, and a fight scene, and contains both real and fantasy threats. The recognition task contained questions relating to multiple characters and scenes. Each scene could be conceptualised as a separate stimulus in this case. Furthermore, content or thematic analysis could be carried out on data collected in the recall tasks to identify whether there were any particular themes or characters which seemed most memorable. In retrospect, the methodology of Experiment 1 would have ideally included multiple positive, negative and neutral clips, with a between-subjects design. This would increase the number of participants needed by a substantial amount to achieve the desired level of power.

The stimuli in Experiment 2 included image sets to more distinctly separate the stimulus into a range of image stimuli, rather than one long clip. Although this meant there were multiple images for each valence condition, they were still centred on one particular story line, so did not fully address the limitation of Experiment 1. In addition, a positive and negative story was written for each image set. This means that there were two positive and two negative stories, rather than just one. Because the same image sets were matched to the negative and positive version of each story, there was still only one combined stimulus set with positive images and narrative, and one with negative images and narrative. The other two contained an image set and narrative of incongruent valence, which appeared to lead to a different style of processing. Whilst the processing of congruent and incongruent emotional material is an interesting topic to explore, any replications of Experiment 2 could look to instead, include multiple congruent stories, as with Experiment 1. This limitation was fully addressed by including an emotionally positive, negative and neutral version of each of three different stories in Experiment 3, and 4 BAS DRM lists for each valence condition in Experiments 3 and 4.

General design limitations. Near-ceiling effects were observed in Experiments 3 and 4. In Experiment 3 this was believed to reflect the tasks being too easy, so the upper age limit was lowered from 11 years to 10 years in Experiment 4, with the addition of a distraction task between each presentation phase and recognition task. Changing from an immediate to a delayed recognition task should have prevented rehearsal and lowered discrimination, in a similar way to that seen for recall tasks (Tan & Ward, 2000). In fact, discrimination rates actually increased in Experiment 4, which may be due to the slightly longer presentation duration for each stimulus, allowing a deeper level of encoding (Craik & Lockhart, 1972). The near-ceiling effects did not obscure all possible effects from coming to light, as a valence effect was still observed. However, this effect varied from that observed in Experiment 3. It would be advisable to replicate the study with a shorter presentation time and explore whether these findings held, or whether effects vary with duration.

All experiments were also likely to be underpowered to various degrees. Each was designed to have a minimum of 30 observations per cell (in line with guidelines by

Cohen, 1988). Once rotation order was introduced into the analysis as a fixed effect, this decreased to 18-19 in Experiment 1, 3-6 in Experiment 3, and 15 in Experiment 4. Rotation was not included in Experiment 2 because it would have correlated too highly with congruency; there were 26-28 observations per cell.

Knowledge Gained

This thesis has built on current theory of memory performance for emotional material in children, by drawing a distinction between the conditions under which valence effects in memory can be produced experimentally. Few studies have explored differences in memory preference for positive and negative material in non-clinical child samples; even fewer have included comparisons between positive and neutral stimuli. This has allowed for various models of emotional memory processing to be tested more fully.

Additionally, the use of a DRM paradigm in Experiments 1-4 allowed for signal detection analysis, to take accuracy into account, in addition to response bias. This is an improvement over some existing research which analyses and reports hit and false alarm rates separately, thus not accounting for their interaction (For example, Howe, 2007; Howe et al., 2010). This design allows for comparisons between experiments, within the thesis, and to those already published using standardized units of analysis (such as *d*' or *A*'). Additionally, there is limited report of response bias in children; so these findings may be of particular interest to other researchers. Finally, Experiments 1 and 2 used a ratings response scale. This has provided data which could be used for Type II *d*' prime analysis, which would provide insight into children's meta-memory; an area which has scarcely been researched to date.

There are several tentative findings, which emerge from this thesis and add to the body of literature on emotional memory in children. In respect to valence effects, there was little evidence of a general preference toward or away from emotionally valenced, compared to neutral stimuli. A preference away from negative material was observed, but only in two specific conditions across four experiments. Findings are therefore considered too mixed, and not robust enough to claim an overall valence effect. The preference away from negative items was found only in the experiments using the simplest stimuli, and design. When more complex stimuli, such as narratives were used as part of the experimental design (as a primer to provide mood context, or as the material to be remembered), then negative material tended to be remembered more accurately than positive. There was also a tendency toward children responding positively more often for negative content, followed by positive, and responding negatively more often for neutral content. This finding highlights the importance of using stringent analysis, such as signal detection, rather than more simple analyses such as percent correct, which does not account for response bias.

In addition to valence effects, a different pattern of findings was apparent when two stimulus sets of a different valence were presented simultaneously; richness of information became a more important deciding factor for what material would be best remembered, with the richer, narrative content better recognized, regardless of valence.

These patterns of results do not favour evolutionary models that propose only threatening material is privileged in cognitive processing. Cognitive models and theories such as the Network Affect Model (Bower & Cohen, 1982) and semantic elaboration theory (Kensinger, 2004) are partially supported by the findings of this thesis. Although they propose that all emotionally valenced material is privileged, significant differences in recall and recognition between positively and negatively valenced materials were observed. These differences are likely due to stimuli confounds, such as arousal or novelty. Tighter control of these extraneous variables is therefore advised in future studies to tease apart the contributions of valence and arousal.

Future Directions

As discussed, the experiments in Chapter III were limited by not having a neutral condition included, as well as including a mixture of positively and negatively valenced stimuli in each clip in Experiment 1. Next steps would be to refine the stimuli to ensure each clip was entirely valenced in one direction, to include more than one clip per valence, and to include neutral clips. This would allow the original hypotheses to be tested more successfully.

Likewise, Experiment 2 would ideally be replicated with additional neutral images and narratives and more than one stimulus set per valence condition. The tentative conclusions reached from Experiment 2 were that richness of information source overrides valence of information, when multiple sources are presented simultaneously which include negatively valenced stimuli from at least one source. Adding a manipulation where visual and narrative stimuli were presented separately would allow this conclusion to be further tested.

Experiment 3 found no significant effect of context. Previous studies included an instruction for children to picture themselves as the main character, leading to a congruency effect. Future studies could replicate this design with additional self-referencing instructions.

Finally, Experiment 4 included a shallow and deep encoding task. Upon reflection, the shallow task asked children to rate the spelling of the word, which may have led to children sounding-out the word, and therefore reading it. The deep processing task asked children to rate how good or bad the word made them feel. This added a self-referencing aspect. Future studies could explore shallow encoding tasks which related to a non-verbal aspect of the word, such as colour of font (more typically employed as a shallow processing task), and deep encoding tasks which did not encourage self-referencing (shown specifically to interact with processing of valenced material), such as thinking of a semantically related word. Should effects still not support the hypothesis being tested; other potential mechanisms should be researched.

Conclusion

Results of the meta-analysis support the hypothesis that scary TV is a plausible explanation for increasing levels of anxiety in typically developing children. The effect was smaller than perhaps expected, indicating that some qualitative studies may have over-estimated the negative effects negative TV has on internalizing emotions, such as anxiety, depression, and sleep disorders. Negativity preferences in memory were explored as a potential mechanism between viewing scary TV and internalizing emotions. No reliable link was found between the levels of trait anxiety children reported, and levels of negativity preference. It is possible that a shift occurs when anxiety reaches clinically significant levels, where negativity preferences drastically increase. This would need to be tested by comparing the performance of clinically anxious and non-anxious children in similar experiments.

Despite no reliable link between anxiety and memory preferences, some valence effects on memory were observed. Memory performance was better for negative than positive narrative/ word stimuli; a findings which appeared fairly consistent and reliable. However, this depended on richness of stimuli; this effect was only observed where narratives were included either as the material to be remembered or as context immediately prior to stimulus presentation. When stimuli of positive and negative valence were presented simultaneously, richness of material became a stronger predictor of memory than valence, with narratives recognized more accurately than visual items.

There did not appear to be an interaction effect between mood and valence on memory, although mood manipulation was only partially successful. Semantic elaboration was explored as a potential mechanism behind valence effects in memory. However, no positive findings were identified. This could have been due to the specific encoding tasks employed, or might signify that another mechanism has a stronger role in valence effects.

Memory performance generally increases with age, but no firm changes of valence effects across age were identified. Effects may become apparent if using a wider age group, or experiments with higher power. Gender did not appear to moderate memory performance.

Future research should further explore the underlying factors behind response bias, so that conditions can be identified where children may be less likely to provide accurate accounts of emotional events or materials. Memory enhancement for negative material may have particular uses in situations such as eyewitness testimony and during social worker interviews with children. Alternatively, positive memory enhancement may be particularly useful as a tool in the treatment of depression, anxiety and other conditions linked with negative affect. It is therefore imperative that research continues into the precise motivations and mechanisms behind enhancement of emotional memory.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(3), 530–535. doi:10.1016/j.cognition.2013.08.014

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-Point or 5-point Likert-type scale. *Educational and Psychological Measurement, 70*(5), 796–807. doi: 10.1177/0013164410366694

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.).* Arlington, VA: American Psychiatric Publishing.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*(2), 97–123. doi:10.1037/h0033773

Arndt, J. (2013). The influence of forward and backward associative strength on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 747–756. doi:10.1037/a0026375

Aron, A., Aron, E. N., & Coups, E. (2008). *Statistics for Psychology (5th ed).* Upper Saddle River, N.J: Pearson.

Asch, S. E., & Ebenholtz, S. M. (1962). The Principle of Associative Symmetry. *Proceedings of the American Philosophical Society, 106*(2), 135–163.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181–185.

Atkinson, R. C., & Juola, J. F. (1971). Factors influencing speed and accuracy of word recognition (Vol. 94305). Presented at: *Fourth International Symposium on Attention and Performance*. The University of Colorado: Stanford University. Retrieved from https://suppes-corpus.stanford.edu/techreports/IMSSS_177.pdf

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The Psychology of Learning and Motivation* (Vol. 2, pp. 89–195.). New York, NY: Academic Press.

Baayen, R. H., Piepenbrock, R., & van Run, H. (1995). *The CELEX Lexical Database (Version Release 2) [CD-ROM]*. University of Pennsylvania, Philadelphia: Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/ldc96l14

Baddeley, A. (1986). *Working Memory*. Oxford Oxfordshire : New York: OUP Australia and New Zealand.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory* (1st ed.). Hove, England ; New York: Psychology Press.

Baddeley, A .D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639–647.

Baltazar, N. C., Shutts, K., & Kinzler, K. D. (2012). Children show heightened memory for threatening social actions. *Journal of Experimental Child Psychology, 112*(1), 102–110. doi:10.1016/j.jecp.2011.11.003

BARB. (2015). *Weekly Viewing Summary*. Retrieved from http://www.barb.co.uk/whats-new/weekly-viewing-summary

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin, 133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Barrett, P. M., Rapee, R. M., Dadds, M. M., & Ryan, S. M. (1996). Family enhancement of cognitive style in anxious and aggressive children. *Journal of Abnormal Child Psychology, 24*(2), 187–203. doi: 10.1007/BF01441484

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology.* Cambridge University Press.

Bartlett, J. C., Burleson, G., & Santrock, J. W. (1982). Emotional mood and memory in young children. *Journal of Experimental Child Psychology, 34*(1), 59–76. doi:10.1016/0022-0965(82)90031-5

Baumeister, R. F., & Tice, D. M. (1990). Point-counterpoints: Anxiety and social exclusion. *Journal of Social and Clinical Psychology, 9*(2), 165–195. doi: 10.1521/jscp.1990.9.2.165

BBFC (2014). *BBFC Classification Guidelines: Age Ratings You Can Trust.* Retrieved from http://www.bbfc.co.uk/sites/default/files/attachments/BBFC%20Classification%20Guidelines%202014_0.pdf

Beck, A. T. (1985). Theoretical perspectives on clinical anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the Anxiety Disorders* (pp. 183–196). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Becker-Blease, K. A., Finkelhor, D., & Turner, H. (2008). Media exposure predicts children's reactions to crime and terrorism. *Journal of Trauma & Dissociation, 9*, 225–248. doi:10.1080/15299730802048652

Beentjes, J. J., & van der Voort, T. H. A. (1991). Recall and language usage in retellings of televised and printed stories. *Poetics, 20,* 91–104. Doi:10.1016/0304-422X(91)90035-N

Beentjes, J. J., Vooijs, M., & Van Der Voort, T. A. (1993). Children's recall of televised and printed news as a function of test expectation. *Journal of Educational Television, 19*(1), 5–13. doi:10.1080/0260741930190102

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: Developmental issues and implications for DSM-V. *The Psychiatric Clinics of North America, 32*(3), 483–524. doi:10.1016/j.psc.2009.06.002

Bickham, D. S., & Rich, M. (2006). Is television viewing associated with social isolation? Roles of exposure time, viewing context, and violent content. *Archives of Pediatrics & Adolescent Medicine, 160*(4), 387–392. doi:10.1001/archpedi.160.4.387

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73–79. https://doi.org/10.3758/BF03195349

Birleson, P. (1981). The validity of depressive disorder in childhood and the development of a self-rating scale: A research report. *Journal of Child Psychology and Psychiatry*, *22*(1), 73–88. doi:10.1111/j.1469-7610.1981.tb00533.x

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., … Nelson, B. (1996). Childhood and adolescent depression: A review of the past 10 Years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(11), 1427–1439. doi:10.1097/00004583-199611000-00011

Bishop, S. J., Dalgleish, T., & Yule, W. (2004). Memory for emotional stories in high and low depressed children. *Memory, 12*(2), 214–230. doi:10.1080/09658210244000667

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach’s alpha. *BMJ: British Medical Journal, 314*(7080), 572-572. doi:10.1136/bmj.314.7080.572

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin*, *99*(2), 229–246. https://doi.org/10.1037/0033-2909.99.2.229

Bond, L., Carlin, J. B., Thomas, L., Rubin, K., & Patton, G. (2001). Does bullying cause emotional problems? A prospective study of young teenagers. *BMJ, British Medical Journal, 323*(7311), 480–484. doi:10.1136/bmj.323.7311.480

Borenstein. (2009). *Introduction to Meta-Analysis* (1st ed.). Chichester, U.K: John Wiley & Sons.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129–148. doi:10.1037/0003-066X.36.2.129

Bower, G. H., & Cohen, P. R. (1982). Emotional influences in memory and thinking: Data and theory. In M. S. Clark & S. T. Fiske (Eds.), *Affect and Cognition: 17th Annual Carnegie Mellon Symposium on Cognition.* Hillsdale, NJ: Erlbaum.: Psychology Press.

Bower, G. H., Gilligan, S. G., & Monteiro, K. P. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General, 110*(4), 451–473. doi:10.1037/0096-3445.110.4.451

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin, 111*(2), 244–255. doi:10.1037/0033-2909.111.2.244

Bradley, M., & Lang, P. (1999). *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual, and Affective Ratings* (Tech. Rep. No. C-1). *The Center for Research in Psychophysiology, University of Florida.*

Bradley, M., & Lang, P. (1999). *International Affective Digitized Sounds (IADS): Stimuli, instruction manual and affective ratings* (Tech. Rep. No. B-2). *The Center for Research in Psychophysiology, University of Florida.*

Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of memembering phenomenologies: So simple a child can do it. *Child Development, 75*(2), 505–522. doi:10.1111/j.1467-8624.2004.00689.x

Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., & Toglia, M. P. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*(2), 137–154. doi:10.1016/j.jecp.2010.04.013

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: a review of data and theory. *Psychological Bulletin, 134*(3), 343–382. https://doi.org/10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development, 73*(5), 1363–1377.

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science, 19*(9), 919–925. doi:10.1111/j.1467-9280.2008.02177.x

Brainerd, C. J., Toglia, M. P., Reyna, V. F., & Stahl, C. (2008). Emotion and false memory: The Cornell/Cortland norms. Presented at: *The Psychonomic Society*, Chicago.

Braun-Lewensohn, O., Celestin-Westreich, S., Celestin, L.-P., Verte, D., & Ponjaert-Kristoffersen, I. (2009). Adolescents' mental health outcomes according to different types of exposure to ongoing terror Attacks. *Journal of Youth and Adolescence, 38*(6), 850–862. doi:10.1007/s10964-008-9305-8

Broadbent, D. E. (1958). *Perception and Communication.* New York: Oxford University Press.

Brophy, A. L. (1986). Alternatives to a table of criterion values in signal detection theory. *Behavior Research Methods, Instruments, & Computers, 18*(3), 285–286. doi:10.3758/BF03204400

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*(3), 208–216. https://doi.org/10.3758/BF03197718

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology, 50*(1), 419–439. doi:10.1146/annurev.psych.50.1.419

Buchanan, T. W., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In S. C. Moore & M. Oaksford (Eds.), *Emotional cognition: From Brain to Behaviour* (pp. 9–34). Amsterdam, Netherlands: John Benjamins Publishing Company.

Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006). False recognition of emotional word lists in aging and Alzheimer disease. *Cognitive and Behavioral Neurology, 19*(2), 71–78. doi:10.1097/01.wnn.0000213905.49525.d0

Buijzen, M., Walma van der Molen, J. H., & Sondij, P. (2007). Parental mediation of children's emotional responses to a violent news event. *Communication Research, 34*(2), 212 – 230. doi:10.1177/0093650206298070

Buratto, L. G., Matthews, W. J., & Lamberts, K. (2009). When are moving images remembered better? Study–test congruence and the dynamic superiority effect. *The Quarterly Journal of Experimental Psychology, 62*(10), 1896–1903. doi:10.1080/17470210902883263

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience 14,* 365-376. doi:10.1038/nrn3475

Cabeza, R., Kapur, S., Craik, F. I., McIntosh, A. R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: A PET study of episodic memory. *Journal of Cognitive Neuroscience,* 9(2), 254–265. doi:10.1162/jocn.1997.9.2.254

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*(3), 401–423. doi:10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, 1*(1), 3–25. doi:10.1207/s15327957pspr0101_2

Cacioppo, John T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components. *Journal of Personality and Social Psychology, 76*(5), 839–855. doi:10.1037/0022-3514.76.5.839

Cahill, L., & McGaugh, J. L. (1996). Modulation of memory storage. *Current Opinion in Neurobiology, 6*(2), 237–242

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature, 371*(6499), 702–704. doi:10.1038/371702a0

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience*, *20*(19), RC99. Retrieved from http://www.jneurosci.org.ezproxy.sussex.ac.uk/content/20/19/RC99

Cantor, J. (1998a). Children's attraction to violent television programming. In J. H. Goldstein (Ed.), *Why We Watch: The Attractions of Violent Entertainment* (Reprint edition, pp. 88–115). New York: Oxford University Press USA.

Cantor, J. (1998b). *"Mommy, I'm Scared": How TV and Movies Frighten Children and What We Can Do to Protect Them* (1st ed.). New York: Mariner Books.

Cantor, J., & Oliver, M. B. (1996). Developmental differences in responses to horror. In J. B. Weaver & R. Tamborini (Eds.), *Horror Films: Research on Audience Preference and Reactions* (pp. 63–80). Hillsdale, NJ: Erlbaum.

Cantor, J., & Reilly, S. (1982). Adolescents' fright reactions to television and films. *The Journal of Communication*, *32*(1), 87–99.

Cantor, J., & Sparks, G. G. (1984). Children's fear responses to mass media: Testing some piagetian predictions. *Journal of Communication*, *34*(2), 90–103. doi:10.1111/j.1460-2466.1984.tb02162.x

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*(1), 85–95. doi:10.1080/14640747308400325

Cartwright-Hatton, S., McNally, D., Field, A. P., Rust, S., Laskey, B., Dixon, C., … Woodham, A. (2011). A new parenting-based group intervention for young anxious children: Results of a randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(3), 242–251.e6. doi:10.1016/j.jaac.2010.12.015

Casey, R. L., Masuda, M., & Holmes, T. H. (1967). Quantitative study of recall of life events. *Journal of Psychosomatic Research*, *11*(2), 239–247. doi:10.1016/0022-3999(67)90013-X

Center for the Study of Emotion and Attention (CSEA-NIMH). (1995). *The International Affective Picture System: Digitized Photographs.* Gainsville, FL: University of Florida.

Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A., & Kemp, A. H. (2014). Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in Psychiatry*, *5*. https://doi.org/10.3389/fpsyt.2014.00080

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General, 132*(2), 310–324. doi: 10.1037/0096-3445.132.2.310

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods, 19*(2), 211–229. doi:10.1037/a0032968

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. *Language, Cognition and Neuroscience*, *29*(10), 1257–1267. doi:10.1080/23273798.2014.897734

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

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences, 106*(14), 6008–6010. doi:10.1073/pnas.0811884106

Cohl, H. A. (2005). *Are We Scaring Ourselves to Death? How Pessimism, Paranoia, and a Misguided Media are Leading Us toward Disaster*. New York: St Martin’s Press.

Coles, M. E., & Heimberg, R. G. (2002). Memory biases in the anxiety disorders: Current status. *Clinical Psychology Review*, *22*(4), 587–627. doi:10.1016/S0272-7358(01)00113-1

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 535–555). New York: Academic Press.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497–505. doi:10.1080/14640748108400805

Comer, J. S., Furr, J. M., Beidas, R. S., Weiner, C. L., & Kendall, P. C. (2008). Children and terrorism-related news: Training parents in coping and media literacy. *Journal of Consulting and Clinical Psychology, 76*, 568–578. doi:10.1037/0022-006X.76.4.568

Conroy, R., & Salmon, K. (2006). Talking about parts of a past experience: The impact of discussion style and event structure on memory for discussed and nondiscussed information. *Journal of Experimental Child Psychology, 95*(4), 278–297. doi:10.1016/j.jecp.2006.06.001

Cordon, I., Melinder, A., Goodman, G., & Edelstein, R. (2013). Children's and adults' memory for emotional pictures: Examining age-related patterns using the Developmental Affective Photo System. *Journal Of Experimental Child Psychology, 114*(2), 339–356. doi:10.1016/j.jecp.2012.08.004

Cortese, M. J., McCarty, D. P., & Schock, J. (2014). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2014.945096

Cortese, Michael J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*(6), 595–609. doi:10.1080/09658211.2010.493892

Craik, F. I. M. (2014). Effects of distraction on memory and cognition: A commentary. *Frontiers in Psychology, 5:* 841. doi:10.3389/fpsyg.2014.00841

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. doi:10.1037/0096-3445.104.3.268

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. doi:10.1007/BF02310555

Crozier, W. R., & Hostettler, K. (2003). The influence of shyness on children's test performance. *The British Journal of Educational Psychology, 73*(3), 317–328. doi:10.1348/000709903322275858

Czapiński, J. (1982). Positive–negative asymmetry at group and individual level: Further evidence and a new interpretation. *Polish Psychological Bulletin, 13*(2), 153–158.

Daleiden, E. L. (1998). Childhood anxiety and memory functioning: A comparison of systemic and processing accounts. *Journal of Experimental Child Psychology, 68*(3), 216–235. doi:10.1006/jecp.1997.2429

Dalgleish, T., Moradi, A. R., Taghavi, M. R., Neshat-Doost, H. T., & Yule, W. (2001). An experimental investigation of hypervigilance for threat in children and adolescents with post-traumatic stress disorder. *Psychological Medicine, 31*(03), 541–547. doi:10.1017/S0033291701003567

Dalgleish, T., Taghavi, R., Neshat-Doost, H., Moradi, A., Canterbury, R., & Yule, W. (2003). Patterns of processing bias for emotional information across clinical Disorders: A comparison of attention, memory, and prospective cognition in children and adolescents with Depression, Generalized Anxiety, and Posttraumatic Stress Disorder. *Journal of Clinical Child & Adolescent Psychology, 32*(1), 10–21. doi:10.1207/S15374424JCCP3201_02

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods, 43*(2), 468–477. doi:10.3758/s13428-011-0064-1

Davey, G. C. (2014). *Psychopathology: Research, Assessment and Treatment in Clinical Psychology* (2nd Revised edition edition). Hoboken, N.J: John Wiley & Sons.

Davies, J. J., & Gentile, D. A. (2012). Responses to children's media use in families with and without siblings: A family development perspective. *Family Relations, 61*(3), 410–425. doi:10.1111/j.1741-3729.2012.00703.x

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*, 65-70.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22. doi:10.1037/h0046671

DeFleur, M. L., Davenport, L., Cronin, M., & DeFleur, M. (1992). Audience recall of news stories presented by newspaper, computer, television and radio. *Journalism & Mass Communication Quarterly, 69*(4), 1010–1022. doi:10.1177/107769909206900419

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 15*(9), 5870–5878.

Derevensky, J. (1974). What children fear. *McGill Journal of Education / Revue Des Sciences de L'éducation de McGill, 9*(001). Retrieved from http://mje.mcgill.ca/article/view/6951

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review, 70*(1), 80–90. doi:10.1037/h0039515

Dewhurst, S. A., Anderson, R. J., & Knott, L. M. (2012). A gender difference in the false recall of negative words: Women DRM more than men. *Cognition and Emotion, 26*(1), 65–74. doi:10.1080/02699931.2011.553037

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology, 12*(4), 541–551. doi:10.1080/095414400750050222

Diamond, G., & Bachman, J. (1986). High-school seniors and the nuclear threat, 1975–1984: Political and mental health implications of concern and despair. *International Journal of Mental Health, 15*(1-3), 210–241. doi:10.1080/00207411.1986.11449030

Dixon, R. A., & von Eye, A. (1984). Depth of processing and text recall in adulthood. *Journal of Reading Behavior, 16*(2), 109–117. doi:10.1080/10862968409547508

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion, 1*(1), 5–11. doi: 10.1037/1528-3542.1.1.5

Donovan, C. L., & Spence, S. H. (2000). Prevention of childhood anxiety disorders. *Clinical Psychology Review, 20*(4), 509–531.

Dorfman, D. D. (1982). RSCORE II: A computer program for accuracy analysis. In J. A. Swets & R. M. Pickett (Eds.), *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (pp. 212–232). New York: Academic Press Inc.

Dudeney, J., Sharpe, L., & Hunt, C. (2015). Attentional bias towards threatening stimuli in children with anxiety: A meta-analysis. *Clinical Psychology Review, 40*, 66–75. doi:10.1016/j.cpr.2015.05.007

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., … Williamson, P. R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLOS ONE, 3*(8), e3081. doi:10.1371/journal.pone.0003081

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.), 315*(7109), 629–634.

Elliott, P. B. (1964). Tables of d'. In J. A. Swets (Ed.), *Signal Detection and Recognition by Human Observers: Contemporary Readings* (pp. 651–684). Wiley.

Eme, R., & Schmidt, D. (1978). The stability of children's fears. *Child Development, 49*(4), 1277–1279. doi:10.2307/1128779

Eysenck, M. W. (1992). *Anxiety: The Cognitive Perspective*. Hove, UK ; Hillsdale, USA: Psychology Press.

Fairbrother, G., Stuber, J., Galea, S., Fleischman, A., & Pfefferbaum, B. (2003). Posttraumatic stress reactions in New York City children after the September 11, 2001, terrorist attacks. *Ambulatory Pediatrics, 3*(6), 304–311. doi:10.1367/1539-4409(2003)003<0304:PSRINY>2.0.CO;2

Famularo, R., Kinscherff, R., & Fenton, T. (1992). Psychiatric diagnoses of maltreated children: Preliminary findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*(5), 863–867. doi:10.1097/00004583-199209000-00013

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. doi:10.3758/BRM.41.4.1149

FCC. (1997). *The TV parental guidelines.* Retrieved from http://www.tvguidelines.org/resources/TheRatings.pdf

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low Kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543–549. doi:10.1016/0895-4356(90)90158-L

Ferguson, C. J., San Miguel, C., & Hartley, R. D. (2009). A multivariate analysis of youth violence and aggression: The influence of family, peers, depression, and media violence. *The Journal of Pediatrics, 155*(6), 904–908.e3. doi:10.1016/j.jpeds.2009.06.021

Festinger, L. (1957). *A Theory of Cognitive Dissonance.* Stanford: Stanford University Press.

Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, *2*(2), 105–124. doi:10.1207/S15328031US0202_02

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, *10*(4), 444–467. doi:10.1037/1082-989X.10.4.444

Field, A. P. (2006a). The behavioral inhibition system and the verbal information pathway to children’s fears. *Journal of Abnormal Psychology, 115*(4), 742–752. doi:10.1037/0021-843X.115.4.742

Field, A. P. (2006b). Watch out for the beast: Fear information and attentional bias in Cchildren. *Journal of Clinical Child & Adolescent Psychology, 35*(3), 431–439. doi:10.1207/s15374424jccp3503_8

Field, A. (2013). *Discovering Statistics Using IBM SPSS Satistics (4th ed.).* Sage Publications.

Field, A. P. (2015). Dread returns to mega-silly one. *Health Psychology Review, 9*(1), 15–20. doi:10.1080/17437199.2013.879198

Field, A. P., & Lester, K. J. (2010). Is there room for “development” in developmental models of information processing biases to threat in children and adolescents? *Clinical Child and Family Psychology Review, 13*(4), 315–332. doi:10.1007/s10567-010-0078-8

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal procedures on free recall. *Psychonomic Science, 19*(4), 249–250. doi:10.3758/BF03328801

Freeman, P. R. (1973). *Table of d’ and B.* Cambridge, England: Cambridge University Press.

Frijda, N. H. (1986). *The Emotions (Studies in Emotion and Social Interaction*). Cambridge; New York; Melbourne; Paris: Cambridge University Press.

Furnham, A. F., & Gunter, B. (1985). Sex, presentation mode and memory for violent and non-violent news. *Journal of Educational Television, 11*(2), 99–105. doi:10.1080/0260741850110203

Gathercole, S. E. (1998). The Development of Memory. *Journal of Child Psychology and Psychiatry, 39*(1), 3–27. doi:10.1111/1469-7610.00301

Gau, S. S. F., Chong, M. Y., Chen, T. H. H., & Cheng, A. T. A. (2005). A 3-year panel study of mental disorders among adolescents in Taiwan. *The American Journal of Psychiatry, 162*(7), 1344–50.

Gehring, R. E., Toglia, M. P., & Kimble, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory & Cognition, 4*(3), 256–260. doi:10.3758/BF03213172

Gentile, D. A., & Walsh, D. A. (2002). A normative study of family media habits. *Journal of Applied Developmental Psychology, 23*(2), 157–178. doi:10.1016/S0193-3973(02)00102-8

Georgeson, M. (n.d.). Sensitivity and Bias: An introduction to Signal Detection Theory [Pdf]. *Postgraduate research methods course.* University of Birmingham: School of Psychology. Retrieved from http://www.birmingham.ac.uk/Documents/college-les/psych/vision-laboratory/sdtintro.pdf

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. doi:10.3758/BF03201693

Gilissen, R., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & van der Veer, R. (2008). Parent–child relationship, temperament, and physiological reactions to fear-inducing film clips: Further evidence for differential susceptibility. *Journal of Experimental Child Psychology, 99*(3), 182–195. doi:10.1016/j.jecp.2007.06.004

Gilissen, R., Koolstra, C. M., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & van der Veer, R. (2007). Physiological reactions of preschoolers to fear-inducing film clips: Effects of temperamental fearfulness and quality of the parent–child relationship. *Developmental Psychobiology, 49*(2), 187–195. doi:10.1002/dev.20188

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331. doi:10.1016/j.learninstruc.2005.07.001

Ginsburg, G. S., Greca, A. M. L., & Silverman, W. K. (1998). Social anxiety in children with anxiety disorders: Relation with social and emotional functioning. *Journal of Abnormal Child Psychology, 26*(3), 175–185. doi:10.1023/A:1022668101048

Glassner, B. (2000). *The Culture of Fear: Why Americans Are Afraid of the Wrong Things*. Basic Books.

Goldstein, H., Yang, M., Omar, R., Turner, R., & Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 49*(3), 399–412. doi:10.1111/1467-9876.00200

Goodman, G. S., & Aman, C. (1990). Children’s use of anatomically detailed dolls to recount an event. *Child Development, 61*(6), 1859–1871. doi:10.1111/j.1467-8624.1990.tb03570.x

Goodman, G. S., Hirschman, J. E., Hepps, D, & Rudy, L. (1991). Children’s memory for stressful events. *Merrill-Palmer Quarterl*y, 37(1), 109–157. Retrieved from http://www.jstor.org/stable/23087341

Gray, P. (2011). The decline of play and the rise of psychopathology in children and adolescents. *American Journal of Play, 3*(4), 443–463.

Greca, A. M. L., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology, 26*(2), 83–94. doi:10.1023/A:1022684520514

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66*(3), 228–234. doi:10.1037/h0023645

Green, D., M., & Swets, J., A. (1966). *Signal Detection Theory and Psychophysics* (1st ed.). New York: John Wiley & Sons, Inc.

Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of childen and young people in Great Britain, 2004*. Palgrave Macmillan: Office for National Statistics. Retrieved from http://www.hscic.gov.uk/catalogue/PUB06116/ment-heal-chil-youn-peop-gb-2004-rep1.pdf

Gunter, B., Furnham, A. F., & Gietson, G. (1984). Memory for the news as a function of the channel of communication. *Human Learning: Journal of Practical Research & Applications, 3*(4), 265–271.

Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: A comparison of three presentation media. *Media Psychology, 2*(2), 93–118. doi:10.1207/S1532785XMEP0202_1

Habermas, T., Meier, M., & Mukhtar, B. (2009). Are specific emotions narrated differently? *Emotion, 9*(6), 751–762. doi:10.1037/a0018002

Hadwin, J. A., & Field, A. P. (Eds.). (2010). *Information Processing Biases and Anxiety: A Developmental Perspective* (1st ed.). Chichester, UK: Wiley-Blackwell.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(3), 251–266. doi:10.1023/A:1010490917575

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400. doi:10.1016/S1364-6613(00)01707-1

Harrison, K., & Cantor, J. (1999). Tales from the screen: Enduring fright reactions to scary media. *Media Psychology, 1*, 97–116. doi:10.1207/s1532785xmep0102_1

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of*d′*. *Behavior Research Methods, Instruments, & Computers, 27*(1), 46–51. doi:10.3758/BF03203619

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 512–521. doi:10.1016/S0022-5371(74)80003-4

Hawes, J., Frainman, E., Webb, J., Orme, S., & Moore, D. (2009). *Merlin - Complete Series 1 Box Set* [DVD]. 2entertain.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis* (1st ed.). Orlando: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. doi:10.1037/1082-989X.3.4.486

Herlitz, A., Nilsson, L.-G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*(6), 801–811. doi:10.3758/BF03211324

Hilton, E., & Gottwald, R. (2001). Differences in visual and auditory short-term memory. *IU South Bend Undergraduate Research Journal, 4.* Retrieved from https://www.iusb.edu/ugr-journal/static/2001/hilton.php

Himmelweit, H. T., Oppenheim, A. N., & Vince, P. (1958). *Television and the Child: An Empirical Study of the Effect of Television on the Young.* Published for the Nuffield Foundation by the Oxford University Press.

Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition, 16*(2), 120–132. doi:10.3758/BF03213479

Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory & Cognition, 17*(2), 175–185. doi:10.3758/BF03197067

Hoekstra, S. J., Harris, R. J., & Helmick, A. L. (1999). Autobiographical memories about the experience of seeing frightening movies in childhood. *Media Psychology, 1*, 117–140. doi:10.1207/s1532785xmep0102_2

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722–738. doi:10.1016/j.learninstruc.2007.09.013

Hoffner, C. (1993). Children's strategies for coping with stress: Blunting and monitoring. *Motivation and Emotion, 17*(2), 91–106. doi:10.1007/BF00995187

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science, 16*(12), 927–931. doi:10.1111/j.1467-9280.2005.01638.x

Howe, M. L. (2007). Children's emotional false memories. *Psychological Science, 18*(10), 856–860. doi:10.1111/j.1467-9280.2007.01991.x

Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*(1), 58–75. doi:10.1080/09658210903476514

Howe, M. L., & Derbish, M. H. (2010). On the susceptibility of adaptive memory to false memory illusions. *Cognition, 115*(2), 252–267. doi:10.1016/j.cognition.2009.12.016

Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology, 108*(1), 77–95. doi:10.1016/j.jecp.2010.06.009

Howe, M. L., Wimmer, M. C., & Blease, K. (2009). The role of associative strength in children's false memory illusions. *Memory, 17*(1), 8–16. doi:10.1080/09658210802438474

Howitt, D. D., & Cramer, P. D. (2010). *Introduction to Statistics in Psychology (5th ed.)*. Harlow, England; New York: Prentice Hall.

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). *Assessing heterogeneity in meta-analysis: Q statistic or I2 index?* (CHIP Documents Paper No. 19). Center for Health, Intervention, and Prevention: University of Connecticut. Retrieved from: http://digitalcommons.uconn.edu/chip_docs/19/

Hughes, A. A., Hedtke, K. A., & Kendall, P. C. (2008). Family functioning in families of children with anxiety disorders. *Journal of Family Psychology, 22*(2), 325–328. doi:10.1037/0893-3200.22.2.325

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275–292. doi:10.1111/1468-2389.00156

Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people.* National Union of Teachers (NUT). Retrieved from: http://www.youngminds.org.uk/assets/0002/2559/Accountability_text_pages.pdf

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53–69. doi:10.1037/0033-2909.104.1.53

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*(2), 130–136. doi:10.1037/0096-1523.1.2.130

Jellinek, M. S., & Snyder, J. B. (1998). Depression and suicide in children and adolescents. *Pediatrics in Review, 19*(8), 255–264. doi:10.1542/pir.19-8-255

Jou, J. (2008). Recall latencies, confidence, and output positions of true and false memories: implications for recall and metamemory theories. *Journal of Memory and Language, 58*(4), 1049–1064. doi:10.1016/j.jml.2007.12.003

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 115–126. doi:10.1016/S0022-5371(84)90082-3

Kendall, P. C. (1985). Toward a cognitive-behavioral model of child psychopathology and a critique of related interventions. *Journal of Abnormal Child Psychology, 13*(3), 357–372. doi: 10.1007/BF00912722

Kendall, P. C., Brady, E. U., & Verduin, T. L. (2001). Comorbidity in childhood anxiety disorders and treatment outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(7), 787–794. doi:10.1097/00004583-200107000-00013

Kensinger, E. A. (2004). Remembering emotional experiences: the contribution of valence and arousal. *Reviews in the Neurosciences, 15*(4), 241–251.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1*(2), 99–113. doi:10.1177/1754073908100432

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169–1180. doi:10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310–3315. doi:10.1073/pnas.0306408101

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience, 6*(2), 110–126. doi:10.3758/CABN.6.2.110

Kensinger, E. A., & Schacter, D. L. (2008). Neural processes supporting young and older adults' emotional memories. *Journal of Cognitive Neuroscience, 20*(7), 1161–1173. doi:10.1162/jocn.2008.20080

Kishiyama, M. M., & Yonelinas, A. P. (2003). Novelty effects on recollection and familiarity in recognition memory. *Memory & Cognition, 31*(7), 1045–1051. doi:10.3758/BF03196125

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology, 65*(2), 190–193. doi:10.1037/h0040288

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. doi:10.1002/jrsm.35

Korhonen, P., & Lahikainen, A. R. (2008). Recent trends in young children's television-induced fears in Finland. *Journal of Children and Media*, *2*(2), 147–162. doi:10.1080/17482790802078664

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. doi:10.1016/j.cognition.2009.06.007

Kovacs, M., Gatsonis, C., Paulauskas, S. L., & Richards, C. (1989). Depressive disorders in childhood. IV. A longitudinal study of comorbidity with and risk for anxiety disorders. *Archives of General Psychiatry, 46*(9), 776–782.

Kramer, B. M., Kalick, S. M., & Milburn, M. A. (1983). Attitudes toward nuclear weapons and nuclear war: 1945–1982. *Journal of Social Issues, 39*(1), 7–24. doi:10.1111/j.1540-4560.1983.tb00127.x

Kučera, H., & Francis, W. N. (1967). *Computational Analysis of Present-day American English.* Providence: Brown University Press.

Ladouceur, C. D., Dahl, R. E., Williamson, D. E., Birmaher, B., Ryan, N. D., & Casey, B. J. (2005). Altered emotional processing in pediatric anxiety, depression, and comorbid anxiety-depression. *Journal of Abnormal Child Psychology*, *33*(2), 165–177. doi:10.1007/s10802-005-1825-z

Lagattuta, K. H. (2014). Linking past, present, and future: Children’s ability to connect mental states and emotions across time. *Child Development Perspectives*, *8*(2), 90–95. doi:10.1111/cdep.12065

La Greca, A., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (1996). Symptoms of posttraumatic stress in children after Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology, 64*(4), 712–723. doi: 10.1037/0022-006X.64.4.712

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

Lane, D. G. (2010). *False Recall Serial Position Effects* (Degree of Doctor of Philosophy). University of Sasatchewan, Saskatchewan; Canada. Retrieved from https://ecommons.usask.ca/bitstream/handle/10388/etd-02122011-170810/DavidLaneThesisMarch142011ETD.pdf

Lang, A., Dhillon, K., & Dong, Q. (1995). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, *39*(3), 313–327. doi:10.1080/08838159509364309

Leigland, L. A., Schulz, L. E., & Janowsky, J. S. (2004). Age related changes in emotional memory. *Neurobiology of Aging*, *25*(8), 1117–1124. doi:10.1016/j.neurobiolaging.2003.10.015

Lengua, L. J., Long, A. C., Smith, K. I., & Meltzoff, A. N. (2005). Pre-attack symptomatology and temperament as predictors of children's responses to the September 11 terrorist attacks. *Journal of Child Psychology and Psychiatry*, *46*(6), 631–645. doi: 0.1111/j.1469-7610.2004.00378.x

Lewicka, M., Czapinski, J., & Peeters, G. (1992). Positive-negative asymmetry or "when the heart needs a reason." *European Journal of Social Psychology*, *22*(5), 425–425. doi:10.1002/ejsp.2420220502

Lewinsohn, P. M. P. D., Rohde, P. P. D., & Seeley, J. R. M. S. (1995). Adolescent psychopathology: III. The clinical consequences of comorbidity. *Journal of the American Academy of Child*, *34*(4), 510–519. doi:10.1097/00004583-199504000-00018

Lewis, S., & Clarke, M. (2001). Forest plots: trying to see the wood and the trees. BMJ, 322(7300), 1479–1480. doi:10.1136/bmj.322.7300.1479

Lidwell, W., Holden, K., & Butler, J. (2003). *Universal Principles of Design.* Gloucester, Mass: Rockport Publishers.

Littner, Y., Mimouni, F. B., Dollberg, S., & Mandel, D. (2005). Negative results and impact factor: A lesson from neonatology. *Archives of Pediatrics and Adolescent Medicine, 159*, 1036–1037.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36.

Maass, A., & Köhnken, G. (1989). Eyewitness identification: Simulating the "weapon effect." *Law and Human Behavior, 13*(4), 397–408. doi:10.1007/BF01056411

Maccoby, E. E., & Jacklin, C. N. (1974). *The Psychology of Sex Differences.* Stanford University Press.

Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren & C. Lewis (Eds.), *A Handbook for Data Analysis in the Behaviorial Sciences: Methodological Issues* (pp. 21–57). Hillsdale, NJ: Erlbaum.: Psychology Press.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Psychology Press.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199. doi: 10.1037/0033-2909.98.1.185

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430. doi:10.1016/j.concog.2011.09.021

Maratos, E. J., Allan, K., & Rugg, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia, 38*(11), 1452–1465.

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(4), 554–565. doi:10.1097/00004583-199704000-00019

Margolin, G., & Gordis, E. B. (2000). The effects of family and community violence on children. *Annual Review of Psychology, 51*(1), 445–479. doi:10.1146/annurev.psych.51.1.445

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. doi:10.1111/j.2044-8295.1973.tb01322.x

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*(1), 37–49. doi:10.1037/0012-1649.22.1.37

Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills, 112*(2), 331–348. doi:10.2466/03.11.PMS.112.2.331-348

Martin, M., Horder, P., & Jones, G. (1992). Integral bias in naming of phobia-related words. *Cognition & Emotion, 6*(6), 479–486. doi:10.1080/02699939208409700

Mathai, J. (1983). An acute anxiety state in an adolescent precipitated by viewing a horror movie. *Journal of Adolescence, 6*(2), 197–200. doi:10.1016/S0140-1971(83)80027-X

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 2*(1), 33–52. doi:10.1111/j.1745-6916.2007.00028.x

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12*(2), 227–255. doi:10.1016/0272-7358(92)90116-P

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review, 14*(5), 989–993. doi:10.3758/BF03194133

Mazzone, L., Ducci, F., Scoto, M. C., Passaniti, E., D’Arrigo, V. G., & Vitiello, B. (2007). The role of anxiety symptoms in school performance in a community sample of children and adolescents. *BMC Public Health, 7*(1), 347. doi:10.1186/1471-2458-7-347

McDonough, P. (2009). *TV viewing among kids at an eight year high.* Retrieved from The Nielsen Company website: http://www.nielsen.com/us/en/insights/news/2009/tv-viewing-among-kids-at-an-eight-year-high.html

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27,* 1–28. doi:10.1146/annurev.neuro.27.070203.144157

McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*(2), 382–92. doi:10.1037/0278-7393.30.2.382

Medina, J., J. (2008). The biology of recognition memory. *Psychiatric Times, June*, 13–15. Retrieved from http://www.brainrules.net/pdf/JohnMedina_PsychTimes_June08.pdf

Meeter, M., Shohamy, D., & Myers, C. (2009). Acquired equivalence changes stimulus representations. *Journal of the Experimental Analysis of Behavior, 91*(1), 127–141. doi:10.1901/jeab.2009.91-127

Mellor, D., & Moore, K. A. (2013). The use of Likert Scales with children. *Journal of Pediatric Psychology*, *39*(3), 369-379. doi:10.1093/jpepsy/jst079

Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., & Williams, D. (2008). Do children "DRM" like adults? False memory production in children. *Developmental Psychology, 44*(1), 169–181. doi:10.1037/0012-1649.44.1.169

Miller, J. (1996). The sampling distribution of *d*'. *Perception & Psychophysics, 58*(1), 65–72.

Miller, P. J., & Sperry, L. L. (1988). Early talk about the past: The origins of conversational stories of personal experience. *Journal of Child Language, 15*(2), 293–315. doi:10.1017/S0305000900012381

Miller, S. M. (1987). Monitoring and blunting: Validation of a questionnaire to assess styles of information seeking under threat. *Journal of Personality and Social Psychology, 52*(2), 345–353. doi:10.1037/0022-3514.52.2.345

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *The American Journal of Psychology, 112*(1), 113–146.

Mitte, K. (2008). Memory bias for threatening information in anxiety and anxiety disorders: A meta-analytic review. *Psychological Bulletin, 134*(6), 886–911. doi:10.1037/a0013343

Moradi, A. R., Taghavi, R., Neshat-Doost, H. T., Yule, W., & Dalgleish, T. (2000). Memory bias for emotional information in children and adolescents with posttraumatic stress disorder: A preliminary study. *Journal of Anxiety Disorders, 14*(5), 521–534. doi:10.1016/S0887-6185(00)00037-2

Morris, P. E., & Fritz, C. O. (2013). Effect sizes in memory research. *Memory, 21*(7), 832–842. https://doi.org/10.1080/09658211.2013.763984

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 116–133. doi:10.1037/0278-7393.21.1.116

MPAA. (2015*). Film ratings.* Retrieved from http://www.mpaa.org/film-ratings/

Muller, K. E., & Glueck, D. (2016). *Glimmpse (Version 2.2.3).* Denver: University of Colorado. Retrieved from http://glimmpse.samplesizeshop.org/#/

Muris, P., & Field, A. (2008). Distorted cognition and pathological anxiety in children and adolescents. *Cognition & Emotion, 22*(3), 395–421. doi:10.1080/02699930701843450

Muris, P., Huijding, J., Mayer, B., Remmerswaal, D., & Vreden, S. (2009). Ground control to Major Tom: Experimental manipulation of anxiety-related interpretation bias by means of the “space odyssey” paradigm and effects on avoidance tendencies in children. Journal of Anxiety Disorders, 23(3), 333–340. doi:10.1016/j.janxdis.2009.01.004

Muris, P., Merckelbach, H., Mayer, B., & Prins, E. (2000). How serious are common childhood fears? *Behaviour Research and Therapy, 38*(3), 217–228. doi:10.1016/S0005-7967(98)00204-6

Muris, P., Merckelbach, H., Meesters, C., & Van Lier, P. (1997). What do children fear most often? *Journal of Behavior Therapy and Experimental Psychiatry, 28*(4), 263–267. doi:10.1016/S0005-7916(97)00024-4

Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging, 23(*2), 263–286. doi:10.1037/0882-7974.23.2.263

Nasby, W., & Yando, R. (1982). Selective encoding and retrieval of affectively valent information: Two cognitive consequences of children's mood states. *Journal of Personality and Social Psychology, 43*(6), 1244–1253. doi:10.1037//0022-3514.43.6.1244

Näslund, J. C., & Schneider, W. (1991). Longitudinal effects of verbal ability, memory capacity, and phonological awareness on reading performance. *European Journal of Psychology of Education, 6*(4), 375. doi:10.1007/BF03172772

Neshat-Doost, H. T., Taghavi, M. R., Moradi, A. R., Yule, W., & Dalgleish, T. (1998). Memory for emotional trait adjectives in clinically depressed youth. *Journal of Abnormal Psychology, 107*(4), 642–650. doi: 10.1037/0021-843X.107.4.642

Nielson, F. A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microposts. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages.* 93-98 (CEUR Workshop Proceedings; No. 718). Crete: Greece.

Norris, C. J., Gollan, J., Berntson, G. G., & Cacioppo, J. T. (2010). The current status of research on the structure of evaluative space. *Biological Psychology, 84*(3), 422–436. doi:10.1016/j.biopsycho.2010.03.011

Norris, C. J., Larsen, J. T., Crawford, L. E., & Cacioppo, J. T. (2011). Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias. *Journal of Research in Personality, 45*(1), 100–111. doi:10.1016/j.jrp.2010.12.001

NSPCC. (2016). *What Children Are Telling us About Bullying*. London: NSPCC.

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*(3), 411–421. doi:10.1037//0278-7393.28.3.411

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49*(4), 469–483. doi:10.1016/S0749-596X(03)00080-9

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology. General*, *129*(2), 242–261.

Ofcom. (2014). *Children and Parents: Media Use and Attitudes Report* (Research Document). UK: Ofcom. Retrieved from: www.ofcom.org.uk/__data/assets/pdf_file/0027/76266/childrens_2014_report.pdf

Ofcom. (2016). *Television - What is the watershed?* Retrieved 7 November 2016, from: http://ask.ofcom.org.uk/help/television/watershed

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483–522. doi:10.1037/0033-295X.108.3.483

O'Kearny, R., & Dadds, M. R. (2005). Language for emotions in adolescents with externalizing and internalizing disorders. *Development and Psychopathology, 17*(2), 529–548. doi:10.1017/S095457940505025X

Ollendick, T. H., & King, N. J. (1991). Origins of childhood fears: An evaluation of Rachman's theory of fear acquisition. *Behaviour Research and Therapy, 29(*2), 117–123. doi:10.1016/0005-7967(91)90039-6

Ollendick, T. H., Yang, B., King, N. J., Dong, Q., & Akande, A. (1996). Fears in American, Australian, Chinese, and Nigerian Children and Adolescents: A Cross-Cultural Study. Journal of Child Psychology and *Psychiatry, 37*(2), 213–220. doi:10.1111/j.1469-7610.1996.tb01393.x

Ollendick, T. H., & Yule, W. (1990). Depression in British and American children and its relation to anxiety and fear. *Journal of Consulting and Clinical Psychology, 58*(1), 126–129. doi: 10.1037/0022-006X.58.1.126

Ortiz, C. D., Silverman, W. K., Jaccard, J., & La Greca, A. M. (2011). Children's state anxiety in reaction to disaster media cues: A preliminary test of a multivariate model. Psychological Trauma: *Theory, Research, Practice, and Policy, 3*, 157–164. doi:10.1037/a0020098

Otgaar, H., Candel, I., & Merckelbach, H. (2008). Children's false memories: Easier to elicit for a negative than for a neutral event. *Acta Psychologica, 128*(2), 350–354. doi:10.1016/j.actpsy.2008.03.009

Otto, M. W., Henin, A., Hirshfeld-Becker, D. R., Pollack, M. H., Biederman, J., & Rosenbaum, J. F. (2007). Posttraumatic stress disorder symptoms following media exposure to tragic events: Impact of 9/11 on children at risk for anxiety disorders. *Journal of Anxiety Disorders, 21*(7), 888–902. doi:10.1016/j.janxdis.2006.10.008

Paavonen, E. J., Pennonen, M., Roine, M., Valkonen, S., & Lahikainen, A. R. (2006). TV exposure associated with sleep disturbances in 5- to 6-year-old children. *Journal of Sleep Research, 15*(2), 154–161. doi:10.1111/j.1365-2869.2006.00525.x

Paavonen, E. J., Roine, M., Pennonen, M., & Lahikainen, A. R. (2009). Do parental co-viewing and discussions mitigate TV-induced fears in young children? *Child: Care, Health and Development, 35*(6), 773–780. doi:10.1111/j.1365-2214.2009.01009.x

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communication Research, 21*(4), 516–546. doi:10.1177/009365094021004004

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5*(2), 176–206. doi:10.1016/0010-0285(73)90032-7

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningful values for 925 nouns. Journal of Experimental Psychology, 76(1, Pt.2), 1–25. https://doi.org/10.1037/h0025327

ParentPort. (2012). *Revealed: How parents control children's access to adult media.* Retrieved from: http://www.parentport.org.uk/News/survey

Pearce, L. J., & Field, A. P. (2016). The impact of 'Scary' TV and film on children's internalizing emotions: A meta-analysis. *Human Communication Research, 42*(1), 98-121. doi:10.1111/hcre.12069.

Petersen, R. C. (1979). Scopolamine state-dependent memory processes in man. *Psychopharmacology, 64*(3), 309–314. doi:10.1007/BF00427515

Pezdek, K., & Stevens, E. (1984). Children's memory for auditory and visual information on television. *Developmental Psychology, 20*(2), 212–218. doi:10.1037/0012-1649.20.2.212

Pfefferbaum, B., Jeon-Slaughter, H., Pfefferbaum, R. L., Houston, J. B., Rainwater, S. M., & Regens, J. L. (2008). Media coverage and children's reactions to disaster with implications for primary care and public health. *The Journal of the Oklahoma State Medical Association, 101*(12), 312–317.

Pfefferbaum, B., Nixon, S. J., Krug, R. S., Tivis, R. D., Moore, V. L., Brown, J. M., … Gurwitch, R. H. (1999). Clinical needs assessment of middle and high school students following the 1995 Oklahoma City bombing. *The American Journal of Psychiatry, 156*(7), 1069–1074.

Pfefferbaum, B., Nixon, S. J., Tivis, R. D., Doughty, D. E., Pynoos, R. S., Gurwitch, R. H., & Foy, D. W. (2001). Television exposure in children after a terrorist incident. *Psychiatry, 64*(3), 202–211.

Pfefferbaum, B., Seale, T. W., Brandt, E. N., Pfefferbaum, R. L., Doughty, D. E., & Rainwater, S. M. (2003). Media exposure in children one hundred miles from a terrorist bombing. *Annals of Clinical Psychiatry, 15*(1), 1–8. doi:10.3109/10401230309085664

Pfefferbaum, B., Seale, T. W., McDonald, N. B., Brandt, E. N., Jr, Rainwater, S. M., Maynard, B. T., … Miller, P. D. (2000). Posttraumatic stress two years after the Oklahoma City bombing in youths geographically distant from the explosion. *Psychiatry, 63*(4), 358–370.

Phelps, E. A. (2005). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology, 57*(1), 27–53. doi:10.1146/annurev.psych.56.091103.070234

Phelps, E. A., Labar, K. S., Anderson, A. K., O'connor, K. J., Fulbright, R. K., & Spencer, D. D. (1998). Specifying the contributions of the human amygdala to emotional memory: A case study. *Neurocase, 4*(6), 527–540. doi:10.1080/13554799808410645

Phillips, D., Prince, S., & Schiebelhut, L. (2004). Elementary school children's responses 3 months after the September 11 terrorist attacks: A study in Washington, DC. *American Journal of Orthopsychiatry, 74*(4), 509–528. doi:10.1037/0002-9432.74.4.509

Piaget, J., & Cook, M. (1952). *Origin of Intelligence in the Child.* New York, NY: International University Press.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*(1-12), 125–126. doi:10.3758/BF03342823

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22*(3), 561–569. doi:10.1016/j.janxdis.2007.04.006

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M.-E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology, 72*(4), 235–270. doi:10.1006/jecp.1999.2491

Quas, J. A., Rush, E. B., Yim, I. S., Edelstein, R. S., Otgaar, H., & Smeets, T. (2015). Stress and emotional valence effects on children's versus adolescents' true and false memory. Memory, 0(0), 1–12. doi:10.1080/09658211.2015.1045909

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raaijmakers, Q. A. W., Van Hoof, A., Hart, H., 't., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-type scale items: Neutral or missing values? *International Journal of Public Opinion Research, 12*(2), 208–216.

Rapee, R. M., Kennedy, S., Ingram, M., Edwards, S., & Sweeney, L. (2005). Prevention and early intervention of anxiety disorders in inhibited preschool children. *Journal of Consulting and Clinical Psychology, 73*(3), 488–497. doi:10.1037/0022-006X.73.3.488

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks: SAGE Publications, Inc.

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8- to 18-year-olds.* Menlo Park; CA: The Henry J. Kaiser Family Foundation. Retrieved from https://kaiserfamilyfoundation.files.wordpress.com/2013/01/8010.pdf

Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donson (Ed.), *The Development of Mathematical Skills* (pp. 75–110). Hove, England: Psychology Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. doi:10.1037/0278-7393.21.4.803

Romer, D., Jamieson, K. H., & Aday, S. (2003). Television news and the cultivation of fear of crime. *Journal of Communication, 53*(1), 88–104. doi:10.1111/j.1460-2466.2003.tb03007.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. doi:10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research* (Revised edition). Newbury Park: Sage Publications.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.* Wiley.

Roy-Byrne, P. P., Davidson, K. W., Kessler, R. C., Asmundson, G. J. G., Goodwin, R. D., Kubzansky, L., … Stein, M. B. (2008). Anxiety disorders and comorbid medical illness. *General Hospital Psychiatry, 30*(3), 208–225. doi:10.1016/j.genhosppsych.2007.12.006

Russell, W. A., & Jenkins, J. J. (1954). *The Complete Minnesota Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test.* Department of Psychology, University of Minnesota.

Russo, M. F., & Beidel, D. C. (1994). Comorbidity of childhood anxiety and externalizing disorders: Prevalence, associated characteristics, and validation issues. *Clinical Psychology Review, 14*(3), 199–221. doi:10.1016/0272-7358(94)90008-6

Saylor, C., Cowart, B., Lipovsky, J., Jackson, C., & Finch, A. (2003). Media exposure to September 11 - Elementary school students' experiences   and posttraumatic symptoms. *American Behavioral Scientist, 46*(12), 1622–1642. doi:10.1177/0002764203254619

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*(6), 523–542. doi:10.3758/BF03197149

Schredl, M., Anders, A., Hellriegel, S., & Rehm, A. (2008). TV viewing, computer game playing and nightmares in school children. *Dreaming, 18*(2), 69–76. doi:10.1037/1053-0797.18.2.69

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., … Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine*, *345*(20), 1507–1512. doi:10.1056/NEJM200111153452024

Schwab-Stone, M., Fallon, T., Briggs, M., & Crowther, B. (1994). Reliability of diagnostic reporting for children aged 6-11 years: A test-retest study of the Diagnostic Interview Schedule for Children-Revised. *The American Journal of Psychiatry, 151*(7), 1048–1054.

Schwartz, G. E., Weinberger, D. A., & Singer, J. A. (1981). Cardiovascular differentiation of happiness, sadness, anger, and fear following imagery and exercise. *Psychosomatic Medicine. Psychosomatic Medicine, 43*(4), 343–364.

Sedgwick, P. (2013). Meta-analyses: How to read a funnel plot. *BMJ, 346*. doi:10.1136/bmj.f1342

Senate Committee on the Judiciary. (1999). Children, violence, and the media: A report for parents and policy makers. Retrieved from http://www.indiana.edu/~cspc/ressenate.htm

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior, 6*(1), 156–163. doi:10.1016/S0022-5371(67)80067-7

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. doi:10.1177/0956797611417632

Simons, D., & Silveira, W. R. (1994). Post-traumatic stress disorder in children after television programmes. *BMJ : British Medical Journal, 308*(6925), 389–390.

Singer, J. A., & Salovey, P. (1988). Mood and memory: Evaluating the network theory of affect. *Clinical Psychology Review, 8*(2), 211–251. doi:10.1016/0272-7358(88)90060-8

Singer, J. L. (1980). The powers and limitations of television. In P. H. Tannenbaum (Ed.), *The Entertainment Functions of Television* (1st ed., pp. 31–36). Hillsdale, N.J: Lawrence Erlbaum Associates.

Siqueland, L., Kendall, P. C., & Steinberg, L. (1996). Anxiety in children: Perceived family environments and observed family interaction. *Journal of Clinical Child Psychology, 25*(2), 225–237. doi:10.1207/s15374424jccp2502_12

Smith, S. L., & Moyer-Gusé, E. (2006). Children and the War on Iraq: Developmental Differences in Fear Responses to Television News Coverage. *Media Psychology, 8*(3), 213–237. doi:10.1207/s1532785xmep0803_2

Smith, S. L., & Wilson, B. J. (2002). Children's comprehension of and fear reactions to television news. *Media Psychology, 4*(1), 1–26. doi:10.1207/S1532785XMEP0401_01

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*(2), 203–220. doi:10.3758/BF03196157

Snodgrass, J. G., Berger, G. L., & Haydon, M. (1985). *Human Experimental Psychology.* New York: Oxford University Press.

Söderholm, C., Häyry, E., Laine, M., & Karrasch, M. (2013). Valence and arousal ratings for 420 Finnish Nouns by age and gender. *PLOS ONE, 8*(8), e72859. doi:10.1371/journal.pone.0072859

Sparks, G. G., & Cantor, J. (1986). Developmental differences in fright responses to a television program depicting a character transformation. *Journal of Broadcasting & Electronic Media, 30*(3), 309–323. doi:10.1080/08838158609386626

Spence, S. H. (1997). Structure of anxiety symptoms among children: a confirmatory factor-analytic study. *Journal of Abnormal Psychology, 106*(2), 280–297. doi:10.1037/0021-843X.106.2.280

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy, 36*(5), 545–566. doi:10.1016/S0005-7967(98)00034-5

Spielberger, C. D. (1973). *Manual for the State-Trait Anxiety Inventory for Children.* Palo Alto, CA: Consulting Psychologists Press.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology, 25*(2), 207–222. doi:10.1080/14640747308400340

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. doi:10.3758/BF03207704

Stauffer, J., Frost, R., & Rybolt, W. (1980). Recall and comprehension of radio news in Kenya. *Journalism & Mass Communication Quarterly, 57*(4), 612–617. doi:10.1177/107769908005700409

Stauffer, J., Frost, R., & Rybolt, W. (1981). Recall and learning from broadcast news: Is print better? *Journal of Broadcasting, 25*(3), 253–262. doi:10.1080/08838158109386449

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*(4), 413–424. doi:10.1007/BF02352267

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The Funnel Plot. In H. R. R. Co-Chair, A. J. S. Co-Author, & M. B. D. A. L. PI (Eds.), *Publication Bias in Meta-Analysis* (pp. 73–98). John Wiley & Sons, Ltd.

Stokes, J., Pogge, D., Wecksell, B., & Zaccario, M. (2011). Parent–child discrepancies in report of psychopathology: The contributions of response bias and parenting stress. *Journal of Personality Assessment, 93*(5), 527–536. doi:10.1080/00223891.2011.594131

Streicher, L. H., & Bonney, N. L. (1974). Children talk about television. *Journal of Communication, 24*(3), 54–61. doi:10.1111/j.1460-2466.1974.tb00388.x

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121*(3), 371–394.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*(2), 181–198. doi: 10.1037/0033-2909.99.2.181

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics (6th ed.).* Boston: Pearson.

Talmi, D., Luk, B. T. C., McGarry, L. M., & Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *Journal of Memory and Language*, *56*(4), 555–574. doi:10.1016/j.jml.2007.01.002

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751. doi:10.3758/BF03195864

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The Role of Attention and Relatedness in Emotionally Enhanced Memory. Emotion, 7(1), 89–102. doi:10.1037/1528-3542.7.1.89

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1589–1625. doi:10.1037/0278-7393.26.6.1589

Tarullo, L. B., Richardson, D. T., Radke-Yarrow, M., & Martinez, P. E. (1995). Multiple sources in child diagnosis: Parent-child concordance in affectively ill and well families. *Journal of Clinical Child Psychology*, *24*(2), 173–183. doi:10.1207/s15374424jccp2402_5

Theunissen, N. C. M., Vogels, T. G. C., Koopman, H. M., Verrips, G. H. W., Zwinderman, K. a. H., Verloove-Vanhorick, S. P., & Wit, J. M. (1998). The proxy problem: Child report versus parent report in health-related. *Quality of Life Research*, *7*(5), 387–397. doi:10.1023/A:1008801802877

Thomason, M. E., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G. H., & Gabrieli, J. D. E. (2009). Development of spatial and verbal working memory capacity in the human brain. *Journal of Cognitive Neuroscience, 21*(2), 316–332. doi:10.1162/jocn.2008.21028

Thomsen, D. K., Olesen, M. H., Schnieber, A., & Tønnesvang, J. (2014). The emotional content of life stories: Positivity bias and relation to personality. *Cognition & Emotion*, *28*(2), 260–277. doi:10.1080/02699931.2013.815155

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*(2), 233–256. doi:10.1080/741944069

Toglia, M. P., & Battig, W. F. (1978). *Handbook of Semantic Word Norms (Vol. vii).* Oxford, England: Lawrence Erlbaum.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375–424. doi:10.1016/0162-3095(90)90017-Z

Troy, M., & Sroufe, L. A. (1987). Victimization among preschoolers: Role of attachment relationship history. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*(2), 166–172. doi:10.1097/00004583-198703000-00007

Tulving, E., Donaldson, W., & Bower, G. H. (1972). *Organization of Memory*. New Haven, CT: Academic Press.

TV Licensing. (2011). *TeleScope: A focus on the nation's viewing habits from TV Licensing*. Retrieved from http://www.tvlicensing.co.uk/about/media-centre/news/report-reveals-latest-uk-tv-watching-trends-NEWS35

Twenge, J. M. (2000). The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952-1993. *Journal of Personality and Social Psychology, 79*(6), 1007–1021. doi: 10.1037/0022-3514.79.6.1007

Ucros, C. G. (1989). Mood state-dependent memory: A meta-analysis. *Cognition and Emotion, 3*(2), 139–169. doi:10.1080/02699938908408077

Unz, D., Schwab, F., & Winterhoff-Spurk, P. (2008). TV news – The daily horror? Emotional effects of violent television news. *Journal of Media Psychology: Theories, Methods, and Applications, 20*(4), 141–155. doi:10.1027/1864-1105.20.4.141

U.S. Bureau of the Consensus. (1925-1998). *Statistical Abstract of the United States.* Washington, DC: U.S. Government Printing Office.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. doi:10.1037/0033-2909.134.3.383

Valkenburg, P. M., Cantor, J., & Peeters, A. L. (2000). Fright reactions to television. *Communication Research, 27*(1), 82 –99. doi:10.1177/009365000027001004

Van Bergen, P., Wall, J., & Salmon, K. (2015). The good, the bad, and the neutral: The influence of emotional valence on young children's recall. *Journal of Applied Research in Memory and Cognition, 4*(1), 29–35. doi:10.1016/j.jarmac.2014.11.001

van de Mortel, F. T. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

van der Molen, J. H. W., & Bushman, B. J. (2008). Children's direct fright and worry reactions to violence in fiction and news television programs. *The Journal of Pediatrics, 153*(3), 420–424. doi:10.1016/j.jpeds.2008.03.036

Vandewater, E. A., Bickham, D. S., & Lee, J. H. (2006). Time well spent? Relating television use to children's free-time activities. *Pediatrics, 117*(2), e181–191. doi:10.1542/peds.2005-0812

Vassilopoulos, S. P. (2012). Social anxiety and memory biases in middle childhood: A preliminary study. *Hellenic Journal of Psycology, 9*, 114–131.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428–443. doi:10.1037/1082-989X.10.4.428

Vevea, J. L., Clements, N. C., & Hedges, L. V., (1993). Assessing the effects of selection bias on validity data for the General Aptitude Test Battery. *Journal of Applied Psychology, 78*(6), 981-987.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3).

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods, 41*(2), 534–538. doi:10.3758/BRM.41.2.534

Voss, K. E., Stem, D. E. Jr., & Fotopoulos, S. (2000). A comment on the relationship between coefficient alpha and scale characteristics. *Marketing Letters, 11*(2), 177–191.doi:10.1023/A:1008146924781

Vuontela, V., Steenari, M.-R., Carlson, S., Koivisto, J., Fjällberg, M., & Aronen, E. T. (2003). Audiospatial and visuospatial working memory in 6–13 year old school children. *Learning & Memory, 10*(1), 74–81. doi:10.1101/lm.53503

Walma Van Der Molen, J. H., & Van Der Voort, T. H. A. (1997). Children's recall of television and print news: A media comparison study. *Journal of Educational Psychology, 89*(1), 82–91. doi:10.1037/0022-0663.89.1.82

Walma Van Der Molen, J., & Voort, T. (1998). Children's recall of the news: TV news stories compared with three print versions. *Educational Technology Research and Development, 46*(1), 39–52. doi:10.1007/BF02299828

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

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 687–704. doi:10.1016/S0022-5371(79)90397-9

Waugh, N. C. (1961). Free versus serial recall. *Journal of Experimental Psychology, 62*(5), 496–502. doi:10.1037/h0043891

Weems, C. F., Scott, B. G., Banks, D. M., & Graham, R. A. (2012). Is T.V. traumatic for all youths? The role of preexisting posttraumatic-stress symptoms in the link between disaster coverage and stress. *Psychological Science, 23*(11), 1293–1297. doi:10.1177/0956797612446952

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development, 34*(3), 166–76.

Whaley, C. (1978). Word–nonword classification time. *Journal of Verbal Learning & Verbal Behavior*, *17*(2), 143–154. doi:10.1016/S0022-5371(78)90110-X

Wicks, R. H., & Drew, D. G. (1991). Learning from news: Effects of message consistency and medium on recall and inference making. *Journalism & Mass Communication Quarterly, 68*(1-2), 155–164. doi:10.1177/107769909106800116

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1997). *Cognitive Psychology and Emotional Disorders* (2nd ed.). Chichester ; New York: John Wiley & Sons.

Williams, R. (2012). Applied topics: Moderator analyses: Categorical models and Meta-regression. *The Campbell Collaboration*. Retrieved 18 June, 2016, from http://www.campbellcollaboration.org/artman2/uploads/1/Moderator_Analysis_Williams.pdf

Wilson, B. J. (2008). Media and children's aggression, fear, and altruism. *The Future of Children, 18*(1), 87–118.

Wilson, B. J., Martins, N., & Marske, A. L. (2005). Children's and parents' fright reactions to kidnapping stories in the news. *Communication Monographs, 72*(1), 46–70. doi:10.1080/0363775052000342526

Wilson, C. E. (1974). The Effect of Medium on Loss of Information. *Journalism Quarterly, 51*(1), 111-15.

Wilson Van Voorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample size. *Tutorials in Quantitative Methods for Psychology, 3*(2), 43–50.

Wood, J. (2006). Effect of anxiety reduction on children's school performance and social adjustment. *Developmental Psychology, 42*(2), 345–349. doi:10.1037/0012-1649.42.2.345

Wood, W., Wong, F. Y., & Gregory, J. (1991). Effects of media violence on viewers' aggression in unconstrained social interaction. *Psychological Bulletin, 109*(3), 371–383. doi:10.1037/0033-2909.109.3.371

Woodward Jr., A. E., Bjork, R. A., & Jongeward Jr., R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 608–617. doi:10.1016/S0022-5371(73)80040-4

Woodward, L. J., & Fergusson, D. M. (2001). Life course outcomes of young people with anxiety disorders in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(9), 1086–1093. doi:10.1097/00004583-200109000-00018

World Health Organization. (2010). *International Statistical Classification of Diseases and Related Healh Problems.* (10th revision, 4th ed.). Malta: WHO Library.

Xie, W., & Zhang, W. (2016). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, 1–13. https://doi.org/10.3758/s13421-016-0679-7

Zahn–Waxler, C., Klimes–Dougan, B., & Slattery, M. J. (2000). Internalizing problems of childhood and adolescence: Prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology*, *12*, 443-466.

Zucker, H. R. (2010). *Multiple memory systems?: Serial position dependent false memory effects* (Degree of Bachelor of Arts). University of Michigan, Michigan; United States

Appendix

Chapter III – Supplementary Material.

A series of ANCOVAs was conducted for each outcome variable of the recognition task (sensitivity and response bias in Experiments 1 and 2) and recall task (percentage of true and number of false memories recalled in Experiments 1 and 2, and percentage of visual memories in Experiment 2). In the first step, age was added as a potential covariate, gender in the second, and levels of trait anxiety in the third step. Covariates were only retained in subsequent steps if they were significantly related to the outcome variable. All non-significant main effects and interaction effects between the potential covariates and outcome measures are reported below. All significant relationships are reported within the main report.

Experiment 1: Children's True and False Memories of Emotional Fictional TV

Excitement Ratings

Children did not report significantly different ratings of excitement for the positive (*M* = 2.27, *SE* = 0.22) and the negative clip (*M* = 2.49, *SE* = 0.21), *t*(40) = −1.30, *p* = .199, 95% BCa CI [−0.54, 0.12], *d* = 0.20. Excitement ratings after viewing the positive clip did not differ significantly between boys (*M* = 2.27, *SE* = 0.32), and girls (*M* = 2.26, *SE* = 0.30), *t*(39) = 0.02, *p* = .983, 95% BCa CI [−0.86, 0.88], *d* = 0.01. They did not differ significantly for the negative clip between boys (*M* = 2.59, *SE* = 0.24), and girls (*M* = 2.37, *SE* = 0.37), *t*(31.92) = 0.50, *p* = .618, 95% BCa CI [−0.60, 1.12], *d* = 0.16.

Recognition Task: Covariate Analysis

Sensitivity (*d*'). Age did not significantly interact with the effect of clip valence, *F*(1, 34) = 0.35, *p* = .558, η_p^2 = .00, modality, *F*(1, 34) = 1.39, *p* = .247, η_p^2 = .04, or clip valence by modality, *F*(1, 34) = 0.16, *p* = .692, η_p^2= .01. Gender did not interact significantly with clip valence, $F(1, 33) = 0.72$, $p = .403$, $\eta_p^2 = .02$, modality, $F(1, 33) = 3.10$. $p = .088$, $\eta_p^2 = .09$, or with the clip valence by modality interaction, $F(1, 33) = 0.15$, $p = .698$, $\eta_p^2 = .01$. Trait anxiety was not significantly related to *d*' scores, $F(1, 32) = 0.54$, $p = .469$, $\eta_p^2 = .02$, and did not significantly interact with clip valence, $F(1, 32) = 0.52$, $p = .478$, $\eta_p^2 = .02$, modality, $F(1, 32) = 1.48$, $p = .233$, $\eta_p^2 = .04$, or the clip valence by modality interaction, $F(1, 32) = 1.55$, $p = .222$, $\eta_p^2 = .05$.

Response bias (*C*). Age was not significantly related to response bias, $F(1, 34) = 0.00$, $p = .999$, $\eta_p^2 = .00$, and did not interact significantly with clip valence, $F(1, 34) = 0.01$, $p = .905$, $\eta_p^2 = .00$, modality, $F(1, 34) = 2.63$, $p = .114$, $\eta_p^2 = .07$, or the clip valence by modality interaction, $F(1, 34) = 0.48$, $p = .491$. $\eta_p^2 = .01$. Gender was not significantly related to response bias, $F(1, 34) = 0.01$, $p = .924$, $\eta_p^2 = .00$, and did not significantly interact with the main effect of clip valence, $F(1, 34) = 0.15$, $p = .703$, $\eta_p^2 = .00$, modality, $F(1, 34) = 0.03$, $p = .861$, $\eta_p^2 = .00$, or the interaction between clip valence and modality, $F(1, 34) = 2.34$, $p = .136$, $\eta_p^2 = .06$. Trait anxiety was also not significantly related to response bias, $F(1, 34) = 0.90$, $p = .766$, $\eta_p^2 = .00$, nor did it interact significantly with valence, $F(1, 34) = 0.06$, $p = .814$, $\eta_p^2 = .00$, modality, $F(1, 34) = 0.54$, $p = .466$, $\eta_p^2 = .02$, or the interaction between clip valence and modality, $F(1, 34) = 0.90$, $p = .349$, $\eta_p^2 = .03$.

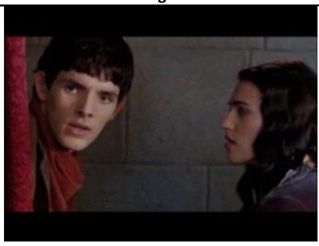
Recall Task: Covariate Analysis

False factual memories. Age did not significantly interact with the effect of clip type on number of false memories recalled, $F(1, 37) = 2.93$, $p = .095$, $\eta_p^2 = .07$. Gender was not significantly related to number of false memories recalled, $F(1, 36) = 4.42$, $p = .461$, $\eta_p^2 = .02$, and did not significantly interact with the effect of clip, $F(1, 36) = 0.39$, $p = .536$, $\eta_p^2 = .01$. Trait anxiety was not significantly related to the number of false memories recalled, *F*(1, 36) = 0.57, *p* = .456, η_p^2 = .02, and did not interact significantly with clip, *F*(1, 36) = 47, *p* = .498, η_p^2 = .01.

True factual memories. Age was not significantly related to the proportion of material recalled, *F*(1, 37) = 2.87, *p* = .099, η_p^2 = .07, and did not significantly interact with the effect of clip, *F*(1, 37) = 0.79, *p* = .381, η_p^2 = .02, or clip by valence, *F*(1, 37) = 0.44, *p* = .510, η_p^2 = .01, on the proportion of material recalled. Gender was not significantly related to the proportion of material recalled, *F*(1, 36) = 0.08, *p* = .776, η_p^2 = .00, and did not interact significantly with the effects of clip, *F*(1, 36) = 2.97, *p* = .093, η_p^2 = .08, valence, *F*(1, 36) = 0.09, *p* = .773, η_p^2 = .00, or clip by valence, *F*(1, 36) = 2.50, *p* = .123, η_p^2 = .07.

Experiment 2: Children's Memories of Emotional Images and Stories

Stimuli: Druid Slides

Image	Positive narrative	Negative narrative
1. 	The Druids are a magical group of people that live outside Camelot. Druids only use magic spells to help people. Some Druids came back to pick up supplies from Camelot. Some of the Druid Children started playing a game where they all hide and the first person home wins. Merlin and Lady Morgana met one of the boys and let him hide in Lady Morgana's room.	The Druids are a magical group of people that live outside Camelot. Some of them have used magic spells to hurt people. King Uther threw them out of the kingdom because magic is not allowed in Camelot. Some Druids came back to cast a spell against King Uther, but one of the Druid boys got separated from his family. Merlin and Lady Morgana did not know the Druids plans and hid him in Lady Morgana's Room.
2. 	There was a gentle knock on the bedroom door. Lady Morgana and Merlin glanced at each other. They knew that if the boy was spotted by his friends he wouldn't win the game. 'Morgana am I allowed in?' Prince Arthur asked from outside her bedroom. "Some children are playing a game and I want to see if there is a Druid boy in here".	There was a loud bang on the bedroom door. Lady Morgana and Merlin looked up in fear. They knew that if the boy was caught he would be put in prison, and maybe even sentenced to death. 'Morgana, open up!' Prince Arthur yelled angrily from outside her bedroom. "I need to come in and search for the Druid Boy!"
3. 	They rushed to hide the boy behind the bed curtain and hoped that Prince Arthur would not spot him. Arthur told them he was searching every room in the castle and did not want to give up until he had found the Druid Boy. He began opening cupboards and drawers for any sign of the boy.	They rushed to hide the boy behind the bed curtain and hoped that Prince Arthur would not spot him. Arthur told them he was searching every room in the castle and did not want to give up until he had found the boy. He began opening cupboards and drawers for any sign of the boy.

4. 	Merlin spotted the boys soft, leather boots on the floor. To keep the boy hidden, Merlin whispered some magical words,and the boots began to walk towards him, all on their own. As Prince Arthur looked round, Merlin told the boots to hide.	Merlin spotted the boys muddy leather boots on the floor. Worried the boy would be caught, Merlin whispered some magical words and the boots began to walk towards him, all on their own. As Prince Arthur looked round, Merlin told the boots to hide.
5. 	Prince Arthur did not see the boots, and left to search the other rooms. Before he left, he told Lady Morgana that if she saw the Druid Boy, she should tell him. He said he would like to win the game. He said that if she told him, he would share his prize with her. But Lady Morgana didn't want to cheat, so she didn't tell him.	Prince Arthur did not see the boots, and left to search the other rooms. Before he left, he told Lady Morgana that if she knew where the Druid Boy was, she had better tell him. He said he would know if she was telling lies. He said if the King found out she was telling lies, he would throw her in prison. But Lady Morgana did not tell him.
6. 	Later that day Merlin, Lady Morgana and Gwen met in the market to discuss how to help the boy get home to win the game. Guards were searching everyone that left the Kingdom so they would have to be very clever to get past them. Merlin offered to take the boy out of Camelot through a secret tunnel that needed a special key. Lady Morgana said she would take the boy because he was her responsibility.	Later that day Merlin, Lady Morgana and Guinevere met in the market to discuss how to help the boy escape from the town. Guards were searching everyone that left the Kingdom so they would have to be very careful to get past them. Merlin offered to take the boy out of Camelot through a dangerous, secret tunnel that needed a special key. Lady Morgana said she would take the boy because he was her responsibility.

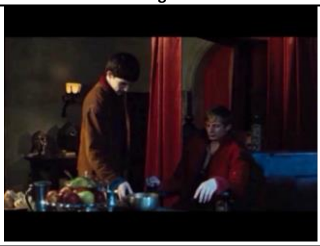
Image	Positive narrative	Negative narrative
7. 	Prince Arthur guarded the key to the secret tunnel out of Camelot. Merlin knew he needed to borrow the key from Prince Arthur or they would never be able to help the boy. Merlin decided his best chance to take the key was when Prince Arthur was eating his dinner. Merlin waited until the evening when he served Arthur some soup.	Prince Arthur guarded the key to the secret tunnel out of Camelot. Merlin knew he must steal the key from Prince Arthur or they would never be able to help the boy. Merlin decided his best chance to take the key was when Prince Arthur was eating dinner. Merlin waited until the evening when he served Arthur some soup.
8. 	Merlin whispered some more magic words. The keys silently slid off the belt around Prince Arthur's waist. They floated up into the air and danced behind Prince Arthur's head as he turned around. It was fun for Merlin to use magic and the trick made him giggle.	Merlin whispered some more magic words. The keys silently slid off the chain around Prince Arthur's waist. They floated up into the air and hid behind Prince Arthur's head as he turned around. It was very dangerous for Merlin to use magic in case he was also caught and put in Prison.
9. 	Arthur could hear the keys jangling but could not spot them. "Can't you hear that noise?" he said to Merlin? Merlin pretended he couldn't hear anything and dropped the keys into the food. Prince Arthur heard the splash and found the keys. He laughed with Merlin at the trick, and Merlin explained the boy had never won the game before so they were helping him.	Arthur could hear the keys jangling but could not spot them. "Can't you hear that noise?" he said to Merlin? Merlin pretended he couldn't hear anything and dropped the keys into the food. Prince Arthur heard the splash and found the keys. He laughed with Merlin at the trick, and Merlin explained the boy had never won the game before so they were helping him.
10. 	Merlin, Prince Arthur and Lady Morgana had a secret meeting about the Druid Boy. They had to keep it secret because they wanted to make sure no-one would spoil the surprise for the boy. Prince Arthur was very good at the game, and knew he would be the best person to help him win. Prince Arthur was very excited to help. He was glad that he and his friends could all join in together.	Merlin, Prince Arthur and Lady Morgana had a secret meeting about the Druid Boy. They had to keep it secret because if anyone found out about it, they would be put in prison. No-one was allowed to help the Druid Boy – not even Prince Arthur. Merlin and Morgana had to try very hard to convince Prince Arthur, because he didn't want to help.
11. 	Prince Arthur agreed to help the Druid Boy get home first. He asked Merlin to meet him and the Druid boy at the end of the tunnel with some tools and a horse. Guards were patrolling the town so he had to be very cunning to get past them. If the guards found Merlin they might also find the Druid Boy and end the game.	Prince Arthur agreed to help the Druid Boy escape from Camelot. He asked Merlin to meet him and the Druid boy at the end of the tunnel with some tools and a horse. Guards were patrolling the town so he had to be very careful not to be caught by them. If the guards found Merlin sneaking about he would surely be thrown in prison.
12. 	The Druid Boy and Prince Arthur went all the way down the tunnel. It was warm and dry, and the floor had straw on it. When they got to the end, they had to wait at the gate. But Prince Arthur and the Druid Boy were still very excited. The Druid Boy knew he would get home and win the game any minute now.	The Druid Boy and Prince Arthur went all the way down the tunnel. It was cold and wet, and it was full of rats. When they got to the end, they couldn't get through the bars. Prince Arthur was angry, and the Druid Boy was very scared. He felt for sure he would never be able to go home again.

Image	Positive narrative	Negative narrative
13. 	Merlin followed their plan and met them exactly on time. They used the key and the tools to open the gate at the end of the tunnel. Arthur had brought the fastest, friendliest horse with him. Merlin and Prince Arthur helped the Druid Boy onto the horse	Merlin was late to meet them at the end of the tunnel. Prince Arthur told him off for being late. They used the key and the tools to open the gate at the end of the tunnel. They thought they would most definitely be caught, and were as quiet as they could be. Merlin and Prince Arthur pulled the Druid Boy onto the horse.
14. 	Prince Arthur and the Druid boy left Camelot. The journey was very quick. They could hear birds tweeting and leaves rustling in the gentle, warm breeze. Prince Arthur was sure they would make it back first and the Druid boy started to feel very excited.	Prince Arthur and the Druid boy left Camelot. They rode for hours and hours. There were creaks and growls coming from all around them. Prince Arthur was sure they were being followed and that somebody would find out what he was doing.
15. 	Soon, they reached the Druid Boy's home. The boy had made it back the fastest and won! As Prince Arthur rode away, the Druids smiled. They were happy to have the boy home and were pleased he had finally won the game.	Eventually they reached the Druid Boy's home. Prince Arthur asked them not to tell anyone he had helped. As Prince Arthur rode away, the Druids smiled. They had discovered a better way into Camelot and could think up a new plan to kill King Uther.

Stimuli: Monster Slides

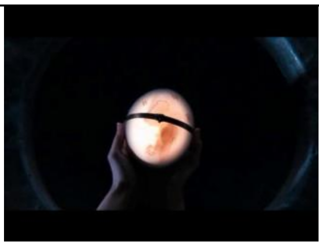
Image	Positive narrative	Negative narrative
1. 	Nimueh is a nice witch who lives in a cave outside of Camelot. She uses magic only for good, to help people the people of Camelot. Nimueh is friends with King Uther and many other people in the kingdom. She watches the people of Camelot to make sure she can keep them safe.	Nimueh is a nasty witch who lives in a cave outside of Camelot. King Uther threw her out of the kingdom because she used magic spells to hurt people. Nimueh is very angry at King Uther and wants everyone in Camelot dead. She watches the people of Camelot and plans how she will kill them.
2. 	One day Nimueh decided she needed some help to look after the people of Camelot so she made an animal out of some clay. She put the clay animal inside a round, white egg and closed it up tightly. Nimueh said some special magic words as she held the egg over the well. The animal inside the egg came to life and began to grow inside the egg.	One day Nimueh decided she wanted to kill all the people of Camelot so she made an animal out of some clay. She put the clay animal inside a round, white egg and locked it up tightly. Nimueh said some evil magic words as she held the egg over the well. The animal inside the egg came to life and began to grow inside the egg.
3. 	Nimueh dropped the egg into the well. It sank down to the bottom of the well. The egg floated through lots of underwater tunnels. It twirled gently round and round as it went.	Nimueh dropped the egg into the well. It plummeted down into the depths of the well. The egg sped through lots of dark, underwater tunnels. It bounced off the jagged walls as it went.

Image	Positive narrative	Negative narrative
4. 	After a short time, the egg reached a pool in another cave. The pool was underneath the Kingdom of Camelot. Once it reached the surface of the water, the shell began to crack and break apart.	After a long time, the egg reached a pool in another cave. The pool was underneath the Kingdom of Camelot. Once it reached the surface of the water, the shell began to crack and break apart.
5. 	Back in Camelot, King Uther found one of his servants asleep on the castle floor. He was was very confused about the sleeping man. “Why is my servant asleep on the floor?” he said to Prince Arthur. “It's day time, and he has a bed of his own.”	Back in Camelot, King Uther found one of his servants dead on the castle floor. He was very angry and upset. “Why has my servant died?” he shouted at Prince Arthur. “Who is going to bring me food and drink now?”
6. 	King Uther called the Doctor, Gaius, to come to the castle and look after the servant. “Is there anything you can do to wake him up?” he asked. Gaius told King Uther that he needed to sleep for a long time to feel better. He suspected there was a bug in Camelot, causing people to fall asleep for a long time.	King Uther called the Doctor, Gaius, to come to the castle and examine the servant. “Is there anything you can do to bring him back to life?” he asked. Gaius told King Uther that he could not, and that the victims died 24 hours after becoming ill. He suspected sorcery and magic might have caused the death.
7. 	Later that day, Gaius found another villager who had caught the bug. He did not know how it was being passed on and wanted to stop it spreading. He asked Merlin to help him take the villager to his house. After thinking very hard, they realised that both people had been drinking the same water in the village.	Later that day, Gaius found another dead body in the village. He was very worried and wanted to find out what was causing people to die. He asked Merlin to help him look at the body in his house. After thinking very hard, they realised both the dead people had been drinking the same water in the village.
8. 	They decided to go down to the cave where the village water comes from. They wanted to collect some of the water to take back with them and test. They wanted to see if it was making people sleepy so they could find a cure. The water pool was not far from the entrance of the cave. Merlin and Gaius soon found it.	They decided to go down to the cave where the village water comes from. They wanted to collect some of the water to take back with them and test. They wanted to see if it was making people sleepy so they could find a cure. The water pool was hidden deep down in the cold, dark cave. Merlin and Gaius had to walk for a long time to get there.
9. 	When they got to the pool Merlin filled up a small bottle with water. They hoped it would help the rest of the villagers. When the bottle of water was safely tucked away, Merlin and Gaius turned to leave the cave. They began to hear a low purring sound from the pool.	When they made it to the pool Merlin filled up a small bottle with water. There were bubbles and ripples in the water and Merlin had been afraid to put his hand into it. When the bottle of water was safely tucked away, Merlin and Gaius turned to leave the cave. They began to hear a low rumbling sound, as if something was growling.

Image	Positive narrative	Negative narrative
 10.	As they looked back towards the pool, an animal popped up out of the water. It was not like any creature Merlin or Gaius has seen before. The magical animal made Merlin and Gaius jump. They ran out of the cave excitedly to tell Prince Arthur what they had discovered in the cave.	As they looked back towards the pool, a huge monster sprang out of the water. It was not like any creature Merlin or Gaius had seen before. The slimy monster had big, sharp claws and roared as loudly as it could. They ran out of the cave in horror to tell Prince Arthur what they had discovered in the cave.
 11.	Merlin decided he would like to know what kind of animal he had seen. He looked in all his books to discover what kind of animal lives in water, and is magical, and could jump high up in the air. He found out about all sorts of animals that were similar, so he still didn't know which kind it was. Merlin decided to go back to the cave and find out for himself.	Merlin knew he should find out what kind of monster it was. He looked in all his books to discover what kind of monster lives in water, and is slimy, and has big, sharp claws. But none of the books could tell him what kind of monster it was, or what would kill it. Merlin was very worried.
 12.	Prince Arthur and Lady Morgana went back to the cave with Merlin to see if they could spot the animal again. They wanted to find out what type of animal he was and how he had got there. They wondered whether he would be able to help them find out what was wrong with the water supply.	Prince Arthur and Lady Morgana went back to the cave with Merlin to search for the monster. Merlin thought the Monster was poisoning the drinking water. They wanted to fight and kill the monster to make sure the people of Camelot were safe from it.
 13.	As Prince Arthur turned a corner he spotted the animal. The animal bounded towards him playfully, stopping a few steps away from Prince Arthur. Prince Arthur asked the animal his name, and what he was doing in the cave. "Boben" said the animal. "Nimueh asked me to help her protect the people of Camelot".	As Prince Arthur turned a corner he came face to face with the Monster. The Monster raced towards him, snarling and foaming at the mouth. Prince Arthur jumped out of the way and swung his sword at the Monster. The Monster was too fast and darted into the shadows.
 14.	Merlin and Lady Morgana rushed over to Prince Arthur when they heard him talking to Boben. Prince Arthur introduced them to each other. Boben explained to them that he had been investigating the water in the cave. He had worked out what was making the people sleepy and he had found out how to fix it.	Merlin and Lady Morgana rushed over to Prince Arthur when they heard the monster. They checked whether Prince Arthur had been hurt. He had some scratches but was not badly injured. They spent the rest of the day and all night looking for the monster. They couldn't find him anywhere!
 15.	Merlin, Prince Arthur and Lady Morgana thanked Boben warmly and returned to Camelot. Nimueh cast another spell to clean the village's water in the pool and nobody else caught the bug. Nimueh and Boben continued to watch over Camelot and keep the villagers safe.	Merlin, Prince Arthur and Lady Morgana gave up trying to find the monster and returned to Camelot. The monster was left free to roam around the pool and continued to poison the water. Nimueh and the monster killed many more people in Camelot.

Manipulation Checks: Emotion Ratings of Stimuli

Table 1 presents the findings of a mixed ANOVA exploring the effects of story, modality, and congruency on internalization ratings. Internalization differed significantly between the druid (*M* = 0.65, *SE* = 0.09) and the monster story conditions (*M* = 0.80, *SE* = 0.08), but did not differ by modality or congruency. No significant interactions were observed.

Table 1.

Results of a mixed ANOVA exploring the effects of story, modality, and congruency on internalization ratings.

Effect	*F*	*df*	*p*	η_p^2
Story	6.29	1, 53	.015	.11
Modality	1.03	1, 53	.314	.02
Congruency	0.33	1, 53	.571	.01
Story × congruency	1.05	1, 53	.310	.02
Story × modality	0.00	1, 53	.951	.00
Modality × congruency	0.02	1, 53	.899	.00
Story × modality × congruency	0.72	1, 53	.401	.01

Table 2 presents the findings of a mixed ANOVA exploring the effects of story, modality, and congruency on happiness ratings. Only the main effects of story condition and modality were significant.

Table 2.

Results of a mixed ANOVA exploring the effects of story, modality, and congruency on happiness ratings.

Effect	*F*	*df*	*p*	η_p^2
Story	11.69	1, 53	.001	.18
Modality	8.80	1, 53	.005	.14
Congruency	1.18	1, 53	.177	.03
Story × congruency	1.00	1, 53	.322	.02
Story × modality	0.01	1, 53	.947	.00
Modality × congruency	0.00	1, 53	.992	.00
Story × modality × congruency	1.05	1, 53	.310	.02

Table 3 presents the findings of a mixed ANOVA exploring the effects of valence, modality, gender, and congruency on excitement ratings. Excitement did not vary depending on any variables entered into the analysis.

Table 3.

Results of a mixed ANOVA exploring the effects of valence, modality, gender, and congruency on excitement ratings.

Effect	*F*	*df*	*p*	η_p^2
Valence	0.00	1, 51	.998	.00
Modality	0.34	1, 51	.558	.01
Gender	0.71	1, 51	.404	.01
Congruency	2.00	1, 51	.164	.04
Valence × gender	2.59	1, 51	.113	.05
Valence × congruency	0.02	1, 51	.903	.00
Valence × modality	0.35	1, 51	.558	.01
Modality × gender	1.36	1, 51	.250	.03
Modality × congruency	0.07	1, 51	.788	.00
Gender × congruency	0.27	1, 51	.604	.01
Valence × gender × congruency	1.30	1, 51	.260	.03
Valence × gender × modality	1.36	1, 51	.250	.03
Valence × modality × congruency	0.06	1, 51	.807	.00
Modality × gender × congruency	0.07	1, 51	.788	.00
Valence × modality × gender × congruency	0.07	1, 51	.788	.00

Recognition Task: Covariate Analysis

Sensitivity (*d*'). Age did not interact significantly with modality, *F*(1, 50) = 2.46, *p* = .123, η_p^2 = .05, or story by modality, *F*(1, 50) = 0.01, *p* = .926, η_p^2 = .00. Gender did not significantly predict *d*' scores, *F*(1, 49) = 0.71, *p* = .403, η_p^2 = .01, and did not significantly interact with story, *F*(1, 49) = 0.62, *p* = .436, η_p^2 = .01, modality, *F*(1, 49) = 1.86, *p* = .179, η_p^2 = .03, or story by modality, *F*(1, 49) = 0.02, *p* = .877, η_p^2 = .00. When trait anxiety scores were added as a covariate, they did not significantly predict *d*' scores, *F*(1, 49) = 0.11, *p* = .743, η_p^2 = .00, and did not significantly interact with story, *F*(1, 49) = 0.37, *p* = .546, η_p^2 = .01, modality, *F*(1, 49) = 2.40, *p* = .128, η_p^2 = .05, or story by modality, *F*(1, 49) = 0.02, *p* = .889, η_p^2 = .00.

Response Bias (*C*). Age was not significantly related to *C* scores, $F(1, 50) = 2.67$, $p = .138$, $\eta_p^2 = .04$, and did not significantly interact with story, $F(1, 50) = 0.63$, $p = .437$, $\eta_p^2 = .01$, modality, $F(1, 50) = 0.48$, $p = .490$, $\eta_p^2 = .01$, or story × modality, $F(1, 50) = 0.05$, $p = .825$, $\eta_p^2 = .00$. Gender was not significantly related to *C* scores, $F(1, 50) = 1.85$, $p = .180$, $\eta_p^2 = .04$, and did not significantly interact with story, $F(1, 50) = 0.81$, $p = .372$, $\eta_p^2 = .02$, modality, $F(1, 50) = 1.33$, $p = .254$, $\eta_p^2 = .03$, or story × modality, $F(1, 50) = 0.10$, $p = .754$, $\eta_p^2 = .00$. Trait anxiety scores were not significantly related to *C* scores, $F(1, 50) = 0.33$, $p = .566$, $\eta_p^2 = .01$, and did not significantly interact with story, $F(1, 50) = 1.49$, $p = .228$, $\eta_p^2 = .03$, modality, $F(1, 50) = 1.78$, $p = .188$, $\eta_p^2 = .03$, or story × modality, $F(1, 50) = 0.97$, $p = .330$, $\eta_p^2 = .02$.

Recall Task: Covariate Analysis

False utterances. Age was not significantly related to number of false memories recalled, $F(1, 50) = 0.59$, $p = .446$, $\eta_p^2 = .01$, and did not interact significantly with story, $F(1, 50) = 0.49$, $p = .486$, $\eta_p^2 = .01$. Gender was not significantly related to number of false memories, $F(1, 50) = 0.01$, $p = .941$, $\eta_p^2 = .00$, and did not interact significantly with story, $F(1, 50) = 1.81$, $p = .185$, $\eta_p^2 = .04$. Trait anxiety was not significantly related to number of false memories recalled, $F(1, 50) = 0.12$, $p = .736$, $\eta_p^2 = .00$, and did not significantly interact with story, $F(1, 50) = 3.69$, $p = .060$, $\eta_p^2 = .01$.

Visual Utterances. Age was not significantly related to proportion of visual memories recalled, $F(1, 51) = 0.16$, $p = .690$, $\eta_p^2 = .00$, and did not interact significantly with story, $F(1, 51) = 2.53$, $p = .118$, $\eta_p^2 = .05$. Gender was not significantly related to proportion of visual memories, $F(1, 51) = 0.05$, $p = .819$, $\eta_p^2 = .00$, and did not interact significantly with story, $F(1, 51) = 0.66$, $p = .419$, $\eta_p^2 = .01$. Trait anxiety was not significantly related to proportion of visual memories recalled, $F(1, 51) = 1.14$, $p =$

.290, η_p^2 = .02, and did not significantly interact with story, F(1, 51) = 0.48, p = .494, η_p^2 = .01.

True Narrative Material. Age did not significantly predict proportion of narrative material recalled correctly, F(1, 51) = 2.85, p = .097, η_p^2 = .02, or interact with story, F(1, 51) = 0.69, p = .412, η_p^2 = .01, valence, F(1, 51) = 1.32, p = .257, η_p^2 = .03, or story × valence, F(1, 51) = 1.62, p = .209, η_p^2 = .03, on number of true narrative utterances recalled. Gender was not significantly related to the proportion of narrative material recalled correctly, F(1, 51) = 1.79, p = .379, η_p^2 = .03, and did not significantly interact with story, F(1, 51) = 0.19, p = .662, η_p^2 = .00, valence, F(1, 51) = 0.24, p = .630, η_p^2 = .01, or story × valence, F(1, 51) = 2.19, p = .145, η_p^2 = .04. Trait anxiety was also not significantly related to the proportion of narrative utterances recalled correctly, F(1, 51) = 0.62, p = .433, η_p^2 = .01, and did not significantly interact with story, F(1, 51) = 2.92, p = .093, η_p^2= .05, valence, F(1, 51) = 1.35, p = .251, η_p^2 = .03, or story × valence, F(1, 51) = 0.07, p = .706, η_p^2 = .00.

Chapters IV, V - Stimuli

Presented Words

Table 4

List pool of words presented in the learning phase of the recognition task in Chapters IV and V.

Positive List A		Positive List B	
SWEET	LOVE	SOFT	MUSIC
Sugar	*Heart*	*Pillow*	*Sound*
Candy	Kiss	Light	Note
Taste	Care	Cotton	Piano
Tooth	Admire	Fur	Radio
Honey	Adore	Fluffy	Band
Chocolate	Friendship	Feather	Melody
Cake	Hug	Downy	Concert
Pie	Kindness	Kitten	Instrument
Neutral List C		Neutral List D	
GREEN	CHAIR	FRUIT	FOOT
Arrow	*Wood*	*Vegetable*	*Toe*
Grass	Table	Apple	Shoe
Blue	Seat	Orange	Kick
Colour	Couch	Pear	Sandals
Red	Desk	Banana	Ankle
Leaf	Sofa	Kiwi	Heel
Arrow	Wood	Citrus	Boot
Clover	Cushion	Ripe	Inch
Negative List E		Negative List F	
CRY	THIEF	ANGER	SICK
Sad	*Robber*	*Fear*	*Cold*
Tears	Steal	Mad	Ill
Tissue	Crook	Hate	Hospital
Sorrow	Burglar	Rage	Vomit
Eyes	Bad	Temper	Doctor
Weep	Jail	Fury	Flu
Sob	Gun	Fight	Fever
Frown	Villain	Hatred	Medicine

Note: Critical lures are written in CAPITALS, associated lures in *italics*. Lures were not presented in the learning phase but were included in the recognition phase of the task

Features of presented words. In order to test the effect of valence on memory, the presented word lists should differ significantly on valence scores, and not differ significantly on any other features. Norms for relevant features were obtained from databases, for as many of the stimuli words as available. The lists were compared using a series of between group ANOVAs. Table 5 shows the differences between the positive, neutral, and negative word lists, across a variety of features.

Table 5

Features of valenced word lists presented in the learning phase of the recognition task in Chapters IV and V.

	List Mean (*SD*)			ANOVA Results			
Word Feature	Positive	Neutral	Negative	df	*F*	*p*	η_p^2
Valence	6.87 (0.79)	5.90 (0.66)	3.28 (1.24)	2, 79	107.00	<.001***	.73
Arousal	4.20 (0.96)	3.55 (0.71)	4.90 (1.08)	2, 79	14.27	<.001***	.27
Subjective Familiarity	556. 91 (40.49)	563.64 (45.14)	529.10 (52.02)	2, 65	3.51	.036*	.10
Imageability	556.73 (78.98)	581.60 (58.15)	523.00 (74.03)	2, 65	3.89	.025*	.11
Age of Acquisition	274.07 (62.47)	253.79 (4.63)	323.38 (62.14)	2, 47	6.14	.004**	.21
Word Frequency (Total)	36.63 (60.76)	40.61 (54.91)	52.33 (82.51)	2, 81	0.41	.662	.01
Written Word Frequency	37.77 (63.33)	42.62 (57.68)	54. 27 (86.41)	2, 81	0.41	.666	.01
Spoken Word Frequency	22.09 (43.27)	14.92 (21.35)	27.58 (49.27)	2, 81	0.71	.494	.02
Subjective Frequency	503.36 (79.12)	487. 08 (53.96)	439.29 (97.88)	2, 37	2.35	.110	.11
Word Length							
Character Length	5.54 (1.90)	4.75 (1.04)	4.89 (1.50)	2, 81	2.13	.126	.05
Syllable Length	1.79 (0.74)	1.50 (0.58)	1.46 (0.58)	2, 81	2.15	.123	.05
Phoneme Length	4.61 (1.81)	3.79 (1.13)	4.07 (1.39)	2, 81	2.25	.112	.05
Neighbourhood Size	5.14 (6.19)	5,25 (5.17)	6.32 (6.78)	2, 81	0.32	.726	.01
Subset Frequency	4009.41 (1431.68)	418.21 (1255.66)	30.91 (1131.84)	2, 36	2.42	.103	.12
Superset Frequency	37.08 (63.27)	25.62 (53.63)	139.82 (55.16)	2, 68	1.29	.283	.04

* *p< .05, ** p < .005, *** p < .001*

Note: Positive *M*s and *SD*s include all available items from Lists A and B, neutral includes all available items from Lists C and D, and Negative includes Lists E and F.

Selection of norms. Valence and Arousal scores were taken from the Warriner, Kuperman, and Brysbaert norms for 13,915 English lemmas (2012). Approximately 20 respondents rated each word from 9 (happy; excited) to 1 (unhappy; calm), for valence, and arousal respectively. There were 1,827 valid responders, aged 16-87 years old, with 60% female respondents, recruited through Amazon Mechanical Turk's crowdsourcing website in the US. Most studies use (Bradley & Lang, 1999) Affective Norms for English Words (ANEW) ratings to assess valence. However, this contains only 1,034 words, including 12 from the negative, 8 from the neutral, and 13 from the positive lists. Ratings of valence are not available for the word lists Sweet, Chair, Fruit, Foot, Angry, Cry, and Thief. In comparison, the Warriner, Kuperman, and Brysbaert norms included 27 of the negative, 27 of the neutral, and all 28 of the positive words presented. It was therefore deemed to give more reliable means of the words lists.

All remaining values were provided through the software *N-Watch* (Davis, 2005). Total word frequency was calculated using CELEX word frequency (per million) measures (Baayen, Piepenbrock, & van Rijn, 1995); number of occurrences in COBUILD/ECT corpus divided by 17.9 (reflecting the 17.9 million words in the corpus). Written and spoken word frequency was calculated from CELEX written and spoken word frequency values; number of occurrences in the COBUILD written and spoken corpus divided by 16.6 and 1.3, respectively (reflecting 16.6, and 1.3 million words in each corpus). CELEX word frequency (Baayen, Piepenbrock, & van Run, 1995) was chosen over Kûcera–Francis count (Kučera & Francis, 1967), since it is based on a sample that is both significantly larger and more representative of everyday reading texts (Davis, 2005). Some researchers have argued that familiarity ratings provide a superior index of the relative frequency of exposure to a word than do objective measures of word frequency. Subjective measures of frequency of exposure were therefore included from a corpus of 2,938 monosyllabic words collected by Balota, Pilotti, and Cortese (2001). Subjective familiarity scores from the MRC database (Coltheart, 1981) were also analysed. They were measured on a 7-point Likert scale of 1 (*very unfamiliar/ never encountered*) to 7 (*very familiar/ encountered several times a day*). Values were multiplied by 100 to create a range from 100-700, in line with other *N-Watch* variable scales.

Age of acquisition from Bristol/ Gilhooly-Logie (Gilhooly & Logie, 1980) was measured on a scale of 1-7, representing years of age. Scores were multiplied by 100 to again create a range from 100-700. The Gilhooly-Logie ratings of age of acquisition (Gilhooly & Logie, 1980) were chosen over the Bird, Franklin, and Howard norms (2001). Despite the Bird et al database containing more recent norms, and norms for 1,988 compared to 1,686 words, a far greater number of presented words were included in the Gilhooly-Logie database (50, compared to 27). The ratio of positive, neutral, and negative presented words with age of acquisition scores was also more similar across the Gilhooly-Logie database.

Imageability (the ease with which a mental image of the word can be formed) from the MRC dataset (Max Coltheart, 1981) was also scaled to a range of 100 to 700, where a higher value indicates greater ease in forming a mental image. Imageability scores were taken from the Bristol/ MRC Psycholinguistic Database (Max Coltheart, 1981), which was formed by merging three smaller databases (Gilhooly & Logie, 1980; Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978). This was chosen over the imageability norms collected by Bird et al (2001) due to a much larger overall database, and more of the presented words being contained in the default vocabulary; 68, compared to 13.

Neighbourhood size is determined by counting the number of words that can be formed by substituting a single letter at any of the letter positions within the string (Coltheart, Davelaar, Jonasson, & Besner, 1977). Subset frequency refers to the frequency of any words (per million) which are embedded within the target word (e.g. *arm* is a subset word embedded within the target word *army*). Superset frequency is the frequency of words (per million) which contain the target word (e.g. target word *arm* is contained within the superset word *army*). Neighbourhood size, subset and superset frequency are calculated by *N-watch*.

Chapter IV – Supplementary Materials.

Effects of Emotional Context and Valence on Children's True and False Memories

Covariate Analysis

Sensitivity: *d*' analysis. A series of four mixed ANCOVAs was run using a 3 × 3 × 6 (Story valence [positive, neutral, negative], DRM list valence [positive, neutral, negative], Rotation order [1, 2, 3, 4, 5, 6]) design with *Bonferroni* corrections, to identify any covariates. Story valence and rotation were included as between participants variables, list valence as a repeated measures variable, and *d*' scores as the dependent variable. Age, gender, and trait anxiety scores, were each entered as covariates in cumulative steps, and retained only where they were significant predictors of *d*' scores. All non-significant results are reported below, whilst significant results are reported within the main results.

Age did not significantly predict *d*' scores, $F(1, 63) = 1.22$, $p = .273$, $\eta_p^2 = .02$, or significantly interact with list valence, $F(2, 126) = 0.02$, $p = .983$, $\eta_p^2 = .00$. Gender did not significantly predict *d*' scores, $F(1, 63) = 0.30$, $p = .585$, $\eta_p^2 = .01$, or interact significantly with list valence, $F(2, 126) = 0.42$, $p = .661$, $\eta_p^2 = .01$. Trait anxiety was not significantly related to *d*' scores, $F(1, 63) = 0.67$, $p = .416$, $\eta_p^2 = .01$, and did not interact significantly with list valence, $F(2, 126) = 1.48$, $p = .232$, $\eta_p^2 = .02$.

Response Bias: Criterion *C* analysis. A series of mixed ANCOVAs with *Bonferroni* adjustments, similar to that for *d*' scores were run, using *C* as the dependent measure. Age did not significantly predict *C* scores, $F(1, 63) = 0.00$, $p = .954$, $\eta_p^2 = .00$, or significantly interact with list valence, $F(2, 126) = 0.79$, $p = .458$, $\eta_p^2 = .02$. Gender did not significantly predict *C* scores, $F(1, 63) = 0.20$, $p = .655$, $\eta_p^2 = .00$, or interact significantly with list valence, $F(2, 126) = 0.59$, $p = .558$, $\eta_p^2 = .01$. Trait anxiety did not significantly predict *C* scores, $F(1, 63) = 1.62$, $p = .209$, $\eta_p^2 = .03$, and did not interact significantly with list valence, *F*(2, 126) = 1.43, *p* = .242, η_p^2 = .02. No potential covariates were retained for subsequent analysis.

Mood and Word List Valence Analysis

Additional analysis was conducted excluding rotation order as a fixed effect, to investigate whether increasing power led to any further significant effects. A 3 × 3 (Story [positive, neutral, negative], Valence of DRM list [positive, neutral, negative]) mixed ANOVA with *Bonferroni* corrections was conducted. There was no significant difference between discrimination scores depending on story condition, *F*(2, 79) = 1.07, *p* = .347, η_p^2 = .03, valence of word list, *F*(2, 158) = 2.43, *p* = .092, η_p^2 = .03, or the interaction between the two, *F*(4, 158) = 0.57, *p* = .683, η_p^2 = .01.

Context Analysis on First Block of Presented Lists

Additional analysis was conducted using recognition data from only the first block of word lists presented, including congruency of the valence of the story and word lists as an independent variable. This was to account for any interference effects which could have been caused by a repeated measures design. A 2 × 3 (Congruency of story and list valence [congruent, incongruent], Valence of DRM list [positive, neutral, negative]) between measures ANOVA with *Bonferroni* corrections was conducted. There was no significant difference between discrimination scores in the congruent context (*M* = 2.79, *SE* = 0.13, 95% BCa CI [2.52, 3.02]) and the incongruent context condition (*M* = 2.77, *SE* = 0.10, 95% BCa CI [2.59, 2.94]), *F*(1, 76) = 0.02, *p* = .881, η_p^2 = .00. There were no significant differences between discrimination of positive (*M* = 2.56, *SE* = 0.13, 95% BCa CI [2.27, 2.83]), neutral (*M* = 2.88, *SE* = 0.15, 95% BCa CI [2.60, 3.14]) and negative valence word lists (*M* = 2.90, *SE* = 0.14, 95% BCa CI [2.58, 3.20]), *F*(2, 76) = 1.96, *p* = .148, η_p^2 = .05. The congruency by valence interaction effect on discrimination was also not significant, *F*(1, 76) = 1.75, *p* = .180, η_p^2 = .04.

Chapter V - Supplementary Material

Explicit Elaboration and Valence Effects on Children's Emotional Memory

For analysis of emotion and spelling ratings of words, *d*' scores, and *C* scores, series' of 2 × 2 × 2 (Word type [emotional, neutral], Valence [positive, negative], Rotation [emotional first, neutral first]) mixed ANCOVAs were conducted. Age was added as a potential covariate in the first step, gender in the second. Covariates were retained only where they significantly predicted the outcome variable or interacted significantly with other independent variables. All non-significant main effects and interaction effects between the potential covariates and outcome measures are reported below. All significant effects are reported within the main report.

Ratings of Words in DRM Lists

Ease of spelling (low elaboration condition). Gender did not significantly predict spelling ratings, $F(1, 54) = 0.04$, $p = .838$, $\eta_p^2 = .00$, or interact significantly with word type, $F(1, 54) = 1.46$, $p = .233$, $\eta_p^2 = .03$, so was not retained as a covariate for further analysis.

Emotion evoked (deep processing condition). Age was not a significant predictor of emotional rating, $F(1, 54) = 0.00$, $p = .977$, $\eta_p^2 = .00$, and did not interact significantly with word type, $F(1, 54) = 0.01$, $p = .907$, $\eta_p^2 = .00$. Gender did not significantly predict emotion ratings, $F(1, 54) = 0.01$, $p = .940$, $\eta_p^2 = .00$, or interact significantly with word type, $F(1, 54) = 0.10$, $p = .758$, $\eta_p^2 = .00$. No covariates were retained for further analysis.

Recognition Task: Covariate Analysis

Sensitivity: *d*' analysis. Gender did not significantly predict *d*' scores, $F(1, 109) = 0.00$, $p = .958$, $\eta_p^2 = .00$, or interact significantly with word type, $F(1, 109) = 0.04$, $p = .834$, $\eta_p^2 = .00$, and was not retained as a covariate in subsequent analysis.

Response Bias: Criterion *C* analysis. Age did not significantly predict response bias, *F*(1, 110) = 0 .02, *p* = .891, η_p^2 = .00, and did not interact significantly with word type, *F*(1, 110) = 0.37, *p* = .546, η_p^2 = .00. Gender did not significantly predict response bias, *F*(1, 110) = 0.51, *p* = .479, η_p^2 = .01, or interact significantly with word type, *F*(1, 110) = 1.26, *p* = .263, η_p^2 = .01. Age and gender were not retained as covariates in subsequent analysis.