

Memory judgements: the contribution of detail and emotion to assessments of believability
and reliability

Lucy V. Justice and Harriet M. J. Smith
Department of Psychology, Nottingham Trent University

Corresponding Author:
Dr. Lucy V. Justice
Department of Psychology
Nottingham Trent University
50 Shakespeare Street
Nottingham
NG1 4FQ

Email: lucy.justice@ntu.ac.uk

Author:
Dr. Harriet M. J. Smith
Department of Psychology
Nottingham Trent University
50 Shakespeare Street
Nottingham
NG1 4FQ

Email: harriet.smith02@ntu.ac.uk

Word count: 7652

MEMORY JUDGEMENTS

Abstract

In legal settings, jury members, police, and legal professionals often have to make judgements about witnesses' or victims' memories of events. Without a scientific understanding of memory, (often erroneous) beliefs are used to make decisions. Evaluation of the literature identified two prevalent beliefs that could influence judgements: 1) memory operates like a video recorder therefore, accounts that are detailed are more believable than those containing vague descriptions, and 2) memories recalled with congruent emotion are more believable than those recalled with incongruent emotion. A 2 (emotionality: emotional, non-emotional) x 2 (detail: high, low) factorial design was generated. In line with previous research, participants made believability judgements (Experiment 1) but uniquely, participants were also asked to judge the reliability of the rememberer's recall (Experiment 2). Self-reported confidence, personality measures, and political orientation were also recorded. Believability judgements did not vary as a function of detail or emotion but detailed accounts were judged as more reliable than vague accounts. Confidence and believability were positively correlated, whereas the confidence-reliability relationship was more complex. Personality and political measures were independent of judgements of both constructs. Our results suggest that believability and reliability are distinct constructs and should be examined as such in future research.

Keywords: credibility judgements, reliability, believability, autobiographical memory, emotional memory

MEMORY JUDGEMENTS

Memory judgements: the contribution of detail and emotion to assessments of believability and reliability

At various stages of the legal process, individuals such as the police, legal professionals, and jurors might have to make judgements about witnesses' or victims' memories of experienced events. Indeed, in many cases, memories constitute the primary, or even sole, evidence, meaning that the outcome of these judgements is pivotally important. A body of literature exists which has examined how judgements of memories are made, however this has focused almost entirely on judgements of believability (Bell & Loftus, 1989; Bollingmo, Wessel, Eilertsen, & Magnussen, 2008; Bollingmo, Wessel, Sandvold, Eilertsen, & Magnussen, 2009; Kaufmann, Drevland, Wessel, Overskeid & Magnussen, 2003; Wessel, Bollingmo, Sonstebj, Nielsen, Eilertsen & Magnussen, 2012). Often conceptualised as credibility, these judgements look to establish whether a given account is truthful, such that it overlays objective reality (see Conway, 2005 for a discussion on truth in memory research) and is without deliberate commission or omission. In this paper, we extend investigation of memory judgements to include assessments of the rememberer's reliability. Based on the assumption that the rememberer is recalling without deliberate deception, we conceptualise reliability as the judgement of the accuracy of a given individual's memory for an event, based on the understanding that memory is fallible and error prone (see "sins" of commission, Schacter, 1999). In other words, there may be circumstances where the truthfulness of an individual's statement is not in doubt, but their ability to recall accurately may be (for discussion of these concepts within a legal setting see Rosenthal, 2002). We are interested in understanding if memory judgement research should extend beyond the reach of believability and should also encompass the measurement and analysis of reliability. In this paper, we examine how judgements of believability and reliability are made. We assess

MEMORY JUDGEMENTS

whether these are independent constructs and investigate the extent to which both types of judgements are influenced by the content of the memory account.

Common-sense beliefs about memory

What might underpin lay people's decisions about the believability and reliability of memories? Research shows that common sense beliefs (Conway, Justice & Morrison, 2014) are central (see Bornstein, 2017, for a review). Such beliefs are based on one's experience, one's understanding of the experience of others, and are possibly shaped by cultural influences such as representations of memory in films, books and television shows. Researchers have focused on the way people understand memory across a variety of populations in a range of contexts, enabling us to draw firm conclusions about the nature of common sense beliefs about memory. For instance, studies have investigated general beliefs about memory across the US (Simons & Chabris, 2011) and Norwegian public (Magnussen et al., 2006), as well as research psychologists, clinical psychologists, hypnotherapists, undergraduates (Ost, Easton, Hope, French & Wright, 2017; Patihis, Ho, Tingen, Lilliefeld & Loftus, 2013), and legal professionals (Benton, Ross, Bradshaw, Thomas & Bradshaw, 2006; Wise & Safer, 2004). Beliefs about memories of childhood sexual abuse (CSA) from the public (McGuire & London, 2017; Wessel, Eilertsen, Langnes, Magnussen, & Melinder, 2016), law students (Ernberg & Landström, 2016) and specialist CSA prosecutors (Ernberg, Tidefors, & Landström, 2016) have received more recent attention. Results from this wide body of research constitute a wealth of evidence and converge towards a similar position: individuals who do not have specific understanding of the scientific underpinnings of memory are likely to hold a wide variety of beliefs. Although these memory beliefs may seem plausible, and some may be supported by contemporary memory science, many are likely not to be. Further, an evaluation of this literature shows two prevalent, inaccurate beliefs. Firstly, that memory operates like a video recorder and as such is a permanent, literal

MEMORY JUDGEMENTS

or near-literal, reliable recording of an event (see Niedźwieńska, Neckar & Baran, 2007 and Ost et al., 2017 for latent construct analyses of memory beliefs). Memories recalled with high detail are therefore judged as being more credible than those recalled with vague detail (Conway et al., 2014). Secondly, the literature reveals that memories recalled with an (perceived) emotionally congruent response (Kaufmann et al., 2003) are necessarily more credible than those with an (perceived) emotionally incongruent response.

Detail, emotion and memory

Turning to beliefs assessing memory permanence, Simons and Chabris (2011) found that 63% of a sample of the US public agreed that memory works like a video camera. In the UK, this figure has been found to be between 65% and 70% for the public, and 62% for police (Conway, Justice & Morrison, 2014). Results from Niedźwieńska et al., (2007), Ost et al., (2016) and Patihis et al., (2014), similarly found that individuals tended to support concepts of memory that relate to its believed permanence and stability. It follows then, that individuals who hold this cluster of beliefs would grant more credibility to memories that are rich in detail than those containing vague descriptions, based on the notion that the contents of a memory mirror the proceedings of an event, and as such, most details should be adequately recalled. Indeed, this pattern of findings has been found within the memory beliefs literature (Ernberg et al., 2016; Magnussen et al., 2012) and has been established experimentally across a series of studies (Bell & Loftus, 1989; Reyes, Thompson & Bower, 1980; Wells & Leippe, 1981). Termed “trivial persuasion” by Bell and Loftus (1989), these studies confirmed the ameliorative effect of the inclusion of detail in an account on credibility and guilt judgements. However, these widely held beliefs contradict over thirty years of memory research; it does not follow that a detailed account is necessarily more credible. The level of detail included in an account is affected not only by whether or not the event has been experienced (as opposed to imagined) (Vrij, 2008), but also contextual factors such as

MEMORY JUDGEMENTS

the passage of time. Memories are psychological representations of an event, and as such, do not “record” information in verbatim or even near-verbatim detail. The human memory system is constructive, information is activated from across disparate neural regions and is compiled in an act of remembering. In other words, each time a memory is recalled it is constructed anew, leading to edits, changes and forgetting; autobiographical memory is fallible, impermanent and malleable (see Conway & Pleydell-Pearce, 2000).

It is often found that the public, police and legal professionals believe that memories recounted with congruent emotional affect are more credible than those recalled with incongruent emotional affect (Ernberg et al., 2016; McGuire & London, 2017; Wessel et al., 2016). Incidentally however, Katz, Paddon, and Barnett (2016) in a review of CSA interviews, found that the majority of CSA disclosure by children is neutral or non-emotional. Nevertheless, further support for the emotion-credibility belief comes from experimental research which has found that emotionally congruent accounts of rape (i.e. those that display negative emotions) are judged as more credible than accounts that show no clear, or non-congruent (i.e. positive) emotion (Bollingmo et al., 2008; Bollingmo et al., 2009; Kaufmann et al., 2003). Indeed, this effect is so ubiquitous that it has been termed the “emotional victim effect” (Ask & Landström, 2010) and is thought to occur due to perceived stereotypes about “appropriate” rape victim behaviour (Wessel et al., 2012). As is evident then, a strong link has been established between perceived emotionally-appropriate/congruent recall and credibility judgements. However, this body of research has largely centered around child witnesses in CSA investigations and female victims of rape; little is known about whether emotionality operates as conduit for credibility outside of these areas of evidence giving and as such, our research will be the first to address this.

It is also important to note that emotionality and detail are not univariate concepts, both characteristics can feature concurrently in accounts; a memory can be both, partly, or not

MEMORY JUDGEMENTS

at all detailed and emotional. Little previous research has investigated the co-occurrence of these characteristics in memory accounts (for an early review see Heuer & Reisberg, 1992). Therefore, one of the aims of this research is to investigate the effect of the inclusion of detail and emotion in a memory account on judgements of both believability and reliability. We hypothesise that an account that is detailed *and* emotionally congruent will receive higher believability judgements than accounts that have a lower amount of detail and are less emotional. However, since the vast majority of previous research assessing memory credibility focuses on the judgement of believability, it is unknown how the manipulation of detail and emotion will affect judgements of reliability. It does seem to follow logically that a similar pattern of responding will be found, such that detailed, and emotionally appropriate accounts will be judged as being more reliable than those with lower detail and/or non-emotional content.

Individual differences in beliefs about autobiographical memory

A further concern that has yet to receive attention in research assessing memory judgements is that of individual differences. Previous research exists which suggests that individual differences, specifically personality traits, mediate memory processes (Arana, Meilan & Perez, 2008; Conway & Pleydell-Pearce, 2000; Flehmig, Steinborn, Langner, & Westhoff, 2007; Zhu et al., 2010). Specifically, susceptibility to false memory is associated with high scores of dissociation and low scores of extroversion (Hyman & Billings, 1998; Porter, Birt, Yuille & Lehman, 2000). Further, individuals who score highly in extraversion and low in neuroticism are found to be more accurate witnesses as compared to individuals scoring low on extroversion and high on neuroticism (Areh & Umek, 2004). What is clear then is that there are links between memory and personality traits, however, the majority of this research has focused on the association of personality and memory recall, rather than assessing the link between personality and judgements of the memory of others. To address

MEMORY JUDGEMENTS

this knowledge gap, we employed the Ten Item Personality Inventory (TIPI, Gosling, Rentfrow, & Swann, 2003), which is a validated short form of the Big-5 (Goldberg, 1992) and the Dirty Dozen (Jonason & Webster, 2010), a validated short form of the Dark Triad (Paulhus & Williams, 2002). Since there is no previous research in this area, we speculate that individuals high in agreeableness and openness will be more likely to rate memory accounts as believable, particularly if they contain emotional information.

A further way of capturing the effect of individual differences is to consider how political orientation might influence beliefs about memory. Jost, Glaser, Kruglanski and Sulloway (2003) argue that political conservatism is a belief system underpinned by a need to avoid uncertainty, as well as sensitivity to fear and threat. Behavioural and neurocognitive differences have been found between those who identify as politically conservative or liberal (Amodio, Jost, Master & Yee, 2007; Kanai, Feilden, Firth & Renshaw, 2011). These differences are relevant to the way in which reported memories might be judged. Liberal and conservative individuals exhibit differences in emotional processing, with liberals tending to be more agreeable and therefore more compassionate (Hirsh, DeYoung, Xu & Peterson, 2010). Individuals at either end of the political spectrum may also assess information differently, with evidence pointing towards conservatives being both less tolerant of ambiguity and more dogmatic. A measure of political orientation was included in the study with the aim of assessing whether judgements of memory differed across liberals and conservatives. Whilst there is no previous research in this area, we speculate that individuals who are more politically liberal will be more likely to believe memory accounts, particularly if they are emotional.

The relationship between confidence and memory assessments

Lastly, confidence has been shown to play a major role in courtroom decision making (Wixted & Wells, 2017). In particular, it is consistently found that mock jurors and triers of

MEMORY JUDGEMENTS

fact are more likely to perceive both adult and child witnesses as credible if they appear confident (Cramer, Brodsky & DeCoster, 2009; Leippe, Manion & Romanczyk, 1992). Indeed, it has been found that juries are more likely to convict when witnesses report being confident (Brewer & Burke, 2002; Penrod & Cutler, 1999; Tenney, MacCoun, Spellman & Hastie, 2007). A substantial body of research exists that has examined confidence in one's own memory judgements (Wixted, Mickes, Clark, Gronlund & Roediger, 2015), and other work suggests that jurors confident in their own decisions will fail to weigh evidence presented to them equally, placing more emphasis on the information they believe (Levesque, 2006). However, little research exists that has investigated the relationship between confidence and judgements of other people's memories. As such, to begin understanding this relationship, we took confidence measures of each judgement. We hypothesise that confidence will have a positive association with both judgement of believability and reliability, such that the more believable or reliable a memory is judged to be, the more confident the participant will be in their judgement.

Aims

In sum, we are investigating how two widely held, erroneous beliefs (memory permanence and emotional-credibility) influence judgements of memory believability, but also, uniquely, judgements of rememberer reliability. To do this, we will ask participants to make judgements of vignettes detailing autobiographical memories of a theft of a personal item. This event was chosen since we felt that judgements were less likely to be influenced by the sex of the participant (in comparison to a crime such as rape for example, which may be perceived and judged differently by males and females) and victim responses to a personal theft would likely span emotions that vary in negativity and intensity, affording all vignettes ecological validity. Personality measures (TIPI (Gosling et al., 2003); Dirty Dozen (Jonason & Webster, 2010)) and political orientation of participants will be taken, allowing us to

MEMORY JUDGEMENTS

provide the first investigation into the link between personality and memory judgements.

Finally, we aim to understand if believability and reliability are independent constructs, and if both should be considered in future work assessing the judgements of memory.

Experiment 1: Believability

Previous research investigating how memories are judged has almost entirely focussed on credibility, asking participants to decide whether or not, or how likely it is, that a memory account is truthful (Bell & Loftus, 1989; Bollingmo et al., 2008; Bollingmo et al., 2009; Kaufmann et al., 2003; Wessel et al., 2012). Since there exists a body of literature assessing this measure, we completed the experiment assessing believability first, with the following hypotheses based on previous findings:

1. Accounts high in detail will be judged as more believable than those low in detail.
2. Accounts that include emotion will be judged as more believable than those that are non-emotional.
3. Detail and emotion will have an additive effect, such that highly detailed and emotional accounts will be most believed.
4. Self-rated confidence will be positively correlated with believability judgements across all levels of detail and emotion.

In the absence of previous literature, we also explored the relationships between the following variables:

1. Ratings of agreeableness, openness and believability across accounts varying in detail and emotion.
2. Political orientation and ratings of believability

MEMORY JUDGEMENTS

Methods

Participants. Forty-four participants (29 [65.9%] female), with a mean age of 36.2 years ($SD = 11.9$, range = 21 - 70), took part in the experiment. Participants were recruited via social media (Facebook, Twitter) and through email advertisements at Nottingham Trent University. Participants completed the study individually, online. The project was granted ethical approval from the Nottingham Trent University ethical review panel.

Design. Memory vignettes were randomly presented according to a 2 (detail: high, low) x 2 (emotion: emotional, non-emotional) within-subjects factorial design, resulting in four different memory types 1) low detail, non-emotional, 2) low detail, emotional, 3) high detail, non-emotional, 4) high detail, emotional. To assess consistency of ratings, two experimental blocks were used, meaning that each participant saw each memory type twice, requiring them to rate eight memories. The outcome measures were memory believability and confidence of judgement.

MEMORY JUDGEMENTS

Materials and Procedure. Participants followed an anonymous link to the study. They were provided with an information page detailing key instructions and then asked to provide consent. Demographic information was collected (age and sex) followed by two personality measures: Ten Item Personality Measures (TIPI; (Gosling, Rentfrow, & Swann, 2003)) and the Dirty Dozen (Jonason & Webster, 2010), which together encompass measurements of agreeableness, conscientiousness, extraversion, neuroticism, openness (TIPI) and psychoticism, narcissism and Machiavellianism (Dirty Dozen), to understand if personality traits are associated with judgements of memory believability. Both personality scales were measured on a 7-point scale from 1-strongly disagree to 7-strongly agree. Short versions of scales were used to facilitate a short study duration. Participants were then asked to indicate their political orientation on a 7-point scale from extremely conservative through to extremely liberal. Next, participants were presented with an information screen with the following instructions:

Next you will be presented with eight memories. These memories depict a theft of a personal item. The accounts were originally given to the police following the theft and are now being given to an insurance company to claim back for the stolen item(s). Please read the memories carefully and answer the questions presented after each account.

To ensure that the length of the vignette and the information presentation order was controlled for, all vignettes were designed using the eight elements depicted in Figure 1. Date of event, current activity, criminal event and stolen item were all fixed elements, in that they were not part of the experimental manipulation. For these, eight different yet conceptually similar items were generated, for example, stolen item consisted of gender-neutral personal items such as laptop, car keys or mobile phone.

MEMORY JUDGEMENTS

[INSERT FIGURE 1 ABOUT HERE]

For each of the experimental elements (recall statement, detail: perpetrator, detail: speech and emotionality statement) eight items were generated, four in the low/neutral condition, four in the high. All high and low statements were matched. For example, a low detail was “he was dressed in darkish clothes” and the matched high detail was “he was wearing a dark brown jacket”. Figure 1 shows matched high and low/neutral examples.

Final vignettes were generated by randomly selecting fixed elements, then randomly selecting experimental elements that were condition appropriate. For example, only emotional elements were entered into the pool for emotional vignettes. Some manual selection was then used to ensure that the vignettes were semantically appropriate, and linguistic elements were added to provide the vignette with a narrative-like structure. Appendix 1 lists all vignettes.

Participants were then asked to answer the following questions using the scale: “*In your opinion, is this person telling the truth about what happened?*” and “*How confident are you that your judgement is correct?*” Both questions were rated on an 11-point rating scale presented in steps of 10, ranging from 0 (not at all believable) to 100 (completely believable). Once both these judgements had been made, participants moved on to the next memory vignette. Memories were randomly presented to the participants to prevent any order effects. After completing the rating of all 8 memories, participants were presented with a debrief.

Results

Believability judgements.

[INSERT TABLE 1 ABOUT HERE]

MEMORY JUDGEMENTS

Firstly, consistency of believability rating within memory types was calculated. This analysis was completed to ensure that the manipulations were the main drivers of results, and not the content of the memories themselves.

[INSERT FIGURE 2 ABOUT HERE]

A series of four paired *t*-tests were run for each pair of vignettes. No significant differences were found between believability ratings of high detail, non-emotional ($t(42) = .88, p = .38$), low detail, emotional ($t(43) = 1.01, p = .32$) and low detail, non-emotional ($t(43) = .15, p = .88$) vignettes. However a significant difference was found for vignettes in the high detail, emotional condition ($t(42) = 6.23, p < .001$), see Table 1.

Next, believability ratings were assessed as a function of both detail and emotion. A linear mixed effects model was run using lme4 (Bates, Maechler, Bolker & Walker, 2015) in Rstudio Version 1.1.383 (RStudio Team, 2016) to account for the hierarchical structure of the data (multiple responses from each participant) and to include the random effect of vignette. To obtain *p*-values, models were contrasted with a null model (a model with a constant in place of fixed effects) using likelihood ratio tests. Results revealed that neither the main effect ($\chi^2(2) = 1.70, p = .43$) model nor the interaction model ($\chi^2(3) = 1.73, p = .63$) was significantly different from the null or each other; neither detail, emotion nor the interaction of the two significantly influenced believability ratings see Figure 2.

In light of the significant difference of ratings between high detail and emotional vignettes, we ran an additional model to try and understand what variables might have been affecting believability ratings. We visually assessed the vignettes and noticed that although we controlled for the content, there were differences in overall length. As such, we entered word count along with emotion and detail as an additional predictor in the main effects model. Results revealed that the model was significantly different to the null model ($\chi^2(4) = 11.77, p = .02$). Examination of the model showed that word count was a significant predictor

MEMORY JUDGEMENTS

of believability, highlighting that participants' ratings of believability increased as vignette word count decreased ($b = -.81$).

Individual differences. A linear mixed effects model with participant and vignette as random effects was used to assess believability and confidence ratings as a function of personality and political orientation. Each dimension of the TIPI, the Dirty Dozen and political orientation was entered into the model as a fixed effect, however no significant results were found, showing that none of the personality measures ($\chi^2(5) = 2.93, p = .71$) or political orientation ($\chi^2(8) = 4.61, p = .80$) were associated with believability judgements. Further, neither personality measures ($\chi^2(5) = 6.93, p = .23$) nor political orientation ($\chi^2(7) = 8.31, p = .31$) had a significant effect on confidence ratings.

Confidence. Turning next to confidence ratings, no significant differences were found between confidence ratings of high detail, emotional ($t(42) = -1.95, p = .06$), high detail, non-emotional ($t(42) = .99, p = .33$), low detail, emotional ($t(43) = -.99, p = .33$) and low detail, non-emotional ($t(43) = -.85, p = .40$) vignettes. These results showed that the experimental manipulation yielded similar confidence scores for the vignettes in the same condition. Similar to believability ratings, confidence was modelled as a function of detail and emotion using main effect and interaction linear mixed effects models. Neither model was significantly different from the null model (main effects: $\chi^2(2) = .21, p = .90$, interaction: $\chi^2(3) = .26, p = .97$) showing that confidence ratings were not influenced by the detail or emotionality of the vignettes.

Due to the significant finding of word count in the believability measure, noted above, we also ran a second model with word count as an additional predictor in the main effects model of confidence. However, results revealed no significant difference between the model and the null ($\chi^2(3) = 6.95, p = .07$) indicating that confidence in believability judgement was not influenced by vignette word count.

MEMORY JUDGEMENTS

A significant positive correlation was however found between confidence ratings and believability judgements ($r = 0.42, p < .001$), showing that as believability judgements increase, so too does the associated confidence rating. The correlation between believability and confidence was also investigated across all combinations of detail and emotion to see if the different memory types affected the level of association. Significant positive correlations were found for all four combinations: high detail, emotional ($r = .55, p < .001$), high detail, non-emotional ($r = .42, p < .001$), low detail, emotional ($r = .45, p < .001$) and low detail, non-emotional ($r = .26, p < .02$). These results show that as believability judgements increase so too do confidence ratings.

Discussion

Contrary to our predictions, believability judgements and reported confidence were not affected by detail or emotion, with most vignettes being rated as moderately believable. This finding is surprising given that the existing body of work highlights ubiquitous beliefs held regarding believability, detail and emotion (Benton, Ross, Bradshaw, Thomas & Bradshaw, 2006; Ernberg & Landström, 201; Ernberg, Tidefors, & Landström, 2016; Magnussen et al., 2006; McGuire & London, 2017; Ost, et al., 2017; Patihis et al., 2014; Simons & Chabris, 2011; Wessel, Eilertsen, Langnes, Magnussen, & Melinder, 2016; Wise & Safer, 2004). Based on our results it would appear then that these beliefs do not underlie judgements of believability. Further, contrary to our predictions, no relationship was found between personality type, political orientation and believability judgement. Confidence ratings and believability judgements however, were significantly positively correlated across all vignettes and for each of the four experimental conditions.

In addition, curiously, one vignette (shown below) received significantly higher believability ratings and had a lower standard deviation than all others:

MEMORY JUDGEMENTS

On the 3rd of February, I was going for a walk. He came from nowhere, shoved me over and stole my phone. My memory of the event is very vivid. He had wiry blonde hair and muttered “give me your phone now”. I felt utter panic.

This vignette fell into the high detail, emotional condition, suggesting that this interaction of characteristics gives rise to the highest judgement of believability. However, its counterpart vignette, also high in detail and emotional, was rated the lowest of all vignettes in believability, such that the effect cannot be attributed to the presence of both variables in the high condition. We were interested to understand what might have caused this variability in ratings and after visually inspecting the vignettes noticed a difference in length. Results revealed that word count was a significant predictor of believability ratings with lower word counts predicting higher believability ratings. This finding then suggests that information presented in a succinct manner, regardless of the number of details or the emotionality in the account, is more likely to be believed. Perhaps simplicity of recounting, then, is important in assessing believability. There has been little previous research that has assessed account length and believability ratings, however a study reported in Johnson, Bush and Mitchell (1998) found no difference in believability of vignettes after systematically varying word count. Since these findings are contradictory, and indeed, we did not systematically vary word count, further work is needed here to explore and replicate the effect of word count on believability judgements.

Experiment 2: Reliability

Credibility is broadly referred to as the believability of a rememberer; an assessment of the individual's veracity. However, there may be cases when the authenticity of the individual is not under question, but instead, the reliability of their memory is. Limited research has been conducted around this measurement (for an exception see Reality

MEMORY JUDGEMENTS

Monitoring (Johnson & Raye, 1918)) and so, we present a second experiment which rather than asking for believability judgements, asked participants to judge the reliability of a memory account. This experiment aimed to understand if judgements of a rememberer's reliability could be influenced by the inclusion/exclusion of memory characteristics, specifically, detail and emotion.

Similar to believability judgements, we hypothesised that:

1. Accounts high in detail and will yield increased reliability judgements.
2. Emotional will yield increased reliability judgements.
3. Accounts high in both detail and emotional accounts will yield increased reliability judgements.
4. Confidence will be positively correlated with these judgements across all memory types.

In the absence of previous literature, we also explored the relationships between the following variables

1. Ratings of agreeableness, openness and reliability across accounts varying in detail and emotion
2. Political orientation and ratings of reliability

Method

Participants. Forty participants (28 [68.3%] female), with a mean age of 35.2 ($SD = 12.9$, range = 19 - 71), took part in the experiment. Participants were recruited via social media (Facebook, Twitter) and through email advertisements at Nottingham Trent University. Participants completed the study individually, online.

Design, Materials and Procedure. The design, materials and procedure were identical to those in Experiment 1, apart from the following exception: after each memory was presented, participants were not asked about believability, but instead were asked the

MEMORY JUDGEMENTS

following question: “*In your opinion, does this person have a reliable memory for what happened?*” The response was provided on an 11-point rating scale presented in steps of 10, ranging from 0 (not at all reliable) to 100 (completely reliable).

Results

Reliability judgements.

[INSERT TABLE 2 ABOUT HERE]

Paired *t*-tests were run for each memory type to assess consistency of reliability ratings. No significant differences were found between reliability ratings of high detail, emotional ($t(37) = -.73, p = .42$), high detail, non-emotional ($t(38) = -.30, p = .77$), low detail, emotional ($t(39) = .73, p = .47$) and low detail, non-emotional ($t(39) = -.20, p = .85$) vignettes.

[INSERT FIGURE 3 AND 4 ABOUT HERE]

Next, reliability ratings were assessed as a function of detail and emotion in the vignettes. As in Experiment 1, both main effects and interaction linear mixed effects models were run, and *p*-values were generated by contrasting all models to a null model using likelihood ratio tests. Results showed that the main effects model ($\chi^2(2) = 25.94, p < .001$) and the interaction model ($\chi^2(3) = 26.25, p < .001$) were significantly different from the null. The main effects model and the interaction models were however not significantly different from each other ($\chi^2(1) = .31, p = .58$) showing that the main effects model fit the data best. Paired comparisons with a Tukey correction revealed that the significant results were driven only by the effect of detail, such that vignettes with high detail were rated higher in reliability than those with low detail (mean difference = 16.61, $p < .001$), main effects of detail and emotion are illustrated in Figure 3 and Figure 4 respectively.

Following on from the significant findings of word count and believability ratings in Experiment one, we ran a second linear mixed effects model, including word count as a

MEMORY JUDGEMENTS

predictor alongside detail and emotion. However, results revealed that word count was not a significant predictor of reliability rating ($\chi^2(1) = .27, p = .60$).

Individual Differences. As in Experiment 1, a linear mixed effects model with participant and vignette entered as random effects was used to investigate reliability judgements and confidence ratings as a function of personality and political orientation. For both models, fixed effects were each dimension of the Ten Item Personality Inventory, the Dirty Dozen (Machiavellianism, psychopathy and narcissism) and political orientation. Neither the personality model ($\chi^2(8) = 2.89, p = .94$) or the political orientation model ($\chi^2(5) = 3.31, p = .65$) was significantly different from their respective null model, showing that reliability judgements were not associated with the personality measures taken ($\chi^2(8) = 7.09, p = .53$), or with political orientation ($\chi^2(5) = 5.16, p = .40$).

Confidence. A series of paired *t*-tests revealed that there were no significant differences of confidence ratings for high detail, emotional ($t(37) = -.79, p = .44$), high detail, non-emotional ($t(38) = -.91, p = .37$), low detail, emotional ($t(38) = -.39, p = .70$) and low detail, non-emotional ($t(39) = -.61, p = .54$) vignettes. A linear mixed effects model also showed that there were no significant differences in confidence ratings of reliability judgements between the null model, the main effects model of detail and emotion ($\chi^2(2) = 3.77, p = .15$) or the interaction model ($\chi^2(1) = .22, p = .64$).

We also ran a second model with word count as an additional predictor in the main effects model of confidence. However, results revealed no significant difference between the model and the null ($\chi^2(1) = .62, p = .43$) indicating that confidence in reliability judgement was not influenced by vignette word count.

Finally, a correlation between reliability judgements and confidence ratings was run to assess the relationship between the two measures. Across all memory types, a weak positive correlation was found ($r = .17, p = .005$). Correlations were also run for all four

MEMORY JUDGEMENTS

vignette combinations: high detail, emotional ($r = .36, p = .003$); high detail, non-emotional ($r = .33, p = .005$), low detail, emotional ($r = .05, p = .70$) and low detail, non-emotional ($r = -.14, p = .25$). Assessment of the results indicate that vignettes with high detail, regardless of emotionality yielded significant positive correlations, but this pattern of results was not found for vignettes with low detail, regardless of emotionality.

Discussion

Partially in line with the hypothesis, results revealed that reliability judgements were influenced by the inclusion of detail in a memory account, such that an individual who recalls an account and includes specific detail is more likely to be judged to have a reliable memory than an individual who includes only vague details. However, contrary to the predictions, accounts that were emotional did not receive different reliability judgements than those that were non-emotional. We did however find, unexpectedly, that word count influenced judgements of believability, such that when the vignettes contained fewer words they were rated as more believable. Judgements of believability and reliability were not found to be influenced by personality or political affiliation. As with believability, personality and political measures were independent of reliability judgements. Confidence was found to significantly correlate with believability but only for vignettes in the high detail condition.

General Discussion

In this paper we investigated judgements of autobiographical memories of a theft of a personal item. In particular, we aimed to understand if believability and reliability are distinct constructs, and therefore if they warrant equal and individual examination in future research. Additionally, we were interested in understanding if these judgements were malleable, i.e. if they could be influenced by the content of a memory. An overview of literature indicated that beliefs of memory permanence and emotionality were prevalent amongst the public (and indeed, other groups) and as such, we varied memory accounts across detail (indicative of

MEMORY JUDGEMENTS

permanence) and emotion. Results showed that judgements of believability and reliability were indeed separate constructs, with reliability judgements being influenced by the presence of detail (Experiment 2), and believability judgements uninfluenced by either characteristic (Experiment 1). Personality and political measures were found to be independent of judgements of believability and reliability. Confidence was found to be associated positively with believability judgements, however the relationship between confidence and reliability judgements was more complex, with only vignettes high in detail yielding a significant correlation.

Detail and Emotion

Detail in a memory account influences judgements of reliability, but not believability. It is perhaps the case in the current research that the difference between high and low detail was not enough to elucidate differences in believability judgements, however, the manipulation was strong enough to cause variation in reliability judgements. In the present study it was found that simply adding concrete detail to statements (for instance describing the attacker's jacket as "dark brown" rather than saying that he was wearing "darkish clothes") was enough to raise reliability judgements by an average of nearly two orders on the scale used. There may be a number of reasons as to why vignettes in the high detail condition were rated as more reliable than those in the low condition. Firstly, as has been noted, there exists a pervasive belief (held by around 60% of the population (Conway et al., 2014; Simons et al., 2001) that memory operates like a video camera, recording in verbatim, or near verbatim, the event in question. It is probable then that a high proportion of the participants in this study held this belief too and were more likely to judge a rememberer as reliable if they provided a memory that was in line with this belief, that is, one that is specific and detailed. The belief bolsters the notion that if an event was experienced then it should be remembered, including accurate recollection of detail, regardless of its triviality. Therefore,

MEMORY JUDGEMENTS

vignettes that lacked specificity flout this belief, and shed doubt on the abilities of the rememberer. Further, vignettes that included specific detail may have received high reliability judgements due to perceived confidence of the rememberer (Wixted et al., 2017). A witness whose memory lacks specificity, or contains vague or hazy details, may make them appear to be unconfident or unsure in their testimony, and hence would yield low judgements of reliability. In this study, to afford tight experimental control, we only included additional detail around the perpetrator. Future research could vary where detail is included and investigate the effects. Indeed, as per Bell et al., (1989), trivial detail has been found to have an ameliorative effect on credibility judgements.

Within research assessing beliefs about CSA and rape, a consistent finding is that emotional witnesses are deemed as being more credible than those who are non-emotional (Ask & Landstrom, 2010). In the research presented here, neither believability nor reliability judgements of accounts of a theft were influenced by its emotionality. As we noted previously, what may be important in assessment of credibility may not be the intensity of emotion per se, but the perceived appropriateness of emotion to the event. To this end, the emotional responses included in the study may not have been perceived as emotionally appropriate, or indeed, both responses may have been equally perceived as appropriate, rendering similar ratings of believability and reliability. A further explanation may be that emotion has a heightened effect on the credibility of sexual abuse memories and affects the assessment of other types of memories to a much lesser extent. Indeed, our research was one of the first to assess this belief in memories of crimes that were not of a sexual nature (see also Landstrom, Ask, Sommar & Willen, 2015). Future research should therefore continue to assess the emotional-witness effect (Ask & Landstrom, 2010) across memories of different crimes to assess the extent to which the effect is domain specific. Possibly too, the lack of influence from statement emotionality may be attributable to a lack of context provided with

MEMORY JUDGEMENTS

the memory vignettes. Vignettes in the present study were presented without any demographic information about the rememberer. Perhaps without this information, judgements are attenuated, or are more difficult to make. Indeed, participants may have expectations of emotional appropriateness dependent on both the crime and the victim. In other words, features of the memory account, characteristics of the rememberer, and information about the event may interact to influence judgements. This would be a valuable line of future investigation, having particular pertinence given the onymous nature of many testimonies and statements given in legal settings.

In any case, the present study found that believability and reliability judgements were not influenced by the emotionality of the vignette. The lack of variability in judgements may reflect the simple fact that emotion is not generally used by individuals to judge the believability and reliability of accounts, or at least accounts of a non-sexual nature, as the ones in these experiments were. However, one interesting and unexpected finding was that the word count of the vignette had a significant impact on the ratings of believability, such that an account with fewer words was rated as being more believable than an account with more words. Memories, it seems, are more believable when they are succinct and briefly described. At a first glance, this result seems to be at odds with the findings from our reliability study (Experiment 2) which showed that when more detail is included in an account it is judged to be more reliable. We may ask then, what kind of memory account will be received most favourably? One that is short and concise to improve judgements of believability, or one that is detailed to influence judgements of reliability? It is important, however, to note that what drove judgements of reliability was not the amount of detail, but the clarity and specificity of detail. Since clear details can be expressed concisely, they are also likely to be shorter and hence, judged as more believable. It seems then that detailed accounts that are conveyed concisely are most likely to receive high reliability and high

MEMORY JUDGEMENTS

believability ratings. Nevertheless, this unexpected finding of account length offers a new avenue of research into believability of the memories of others and should be varied systematically in future research.

Personality and Political Measures

Despite the fact that much previous research has found that some personality traits are associated with memory ability / susceptibility (Hyman et al., 1998; Porter et al., 2000), little work has investigated how personality differences affect the judgement of the memories of others, and no studies were found that had assessed the impact of political orientation on memory judgements. Results from the present study however showed neither personality traits (as measured by the TIPI and Dirty Dozen) nor political orientation were significantly associated with believability and reliability judgements or confidence ratings. This suggests that judgements of others' memories are independent of the individual differences tested here.

Judgements and Self-Reported Confidence

Believability judgements showed a consistent relationship with confidence. Across all conditions, confidence ratings and believability judgements were positively correlated, such that higher confidence ratings were associated with higher believability judgements. For reliability and confidence judgements, it appears that only memories in the high detail condition elicited a significant association. This pattern of results shows that the more believable an account is judged to be, the more confident the individual becomes in that judgement; participants therefore had confidence in their own truth detection skills. However, over-confidence in detecting lies has been found in many studies (see DePaulo, Charlton, Cooper, Lindsay & Muhlenbruck, 1997, for a meta-analysis). This finding does not sit neatly with our results, which revealed that lower confidence tended to be expressed when memories were judged to be less believable. However, perhaps in our experiments the status

MEMORY JUDGEMENTS

of the rememberer as an apparent victim, who is therefore vulnerable, has some bearing on confidence. It might simply be the case, then, that the participants were giving the rememberer the “benefit of the doubt” – in other words if they did not judge the memory account to be believable / reliable they were cautious in expressing high confidence.

Limitations and Future Research

We recommend future research on the topic of believability and reliability judgments, both for the purposes of replication, and to develop understanding of these constructs. To support and guide future research, we assess the limitations of our work below.

We ran manipulation checks to ensure that each pair of vignettes with the same levels of detail and emotion were rated similarly, (i.e. we checked that the two vignettes with low detail and low emotion were rated similarly for believability and reliability) and found no significant differences in ratings between all pairs of vignettes across both experiments (with the exception of vignettes high in detail and high in emotion in Experiment 1). Whilst this showed us that our manipulations were successful in that they produced similar ratings, we cannot be certain that the effects were in fact due to the manipulation of detail and emotion since we did not ask participants whether they perceived differences in the content of the vignettes. Future research using vignettes manipulated across a number of variables should ensure that the manipulations are indeed perceived by the participants to ensure that results can be attributed to the variables of interest.

In each experiment we used a within-subjects design to allow for increased statistical power and to reduce the effect of individual differences on results. This design however poses problems as participants will have viewed all eight vignettes and may have attempted to guess the design and / or hypotheses of the experiment which may in turn have affected the results. We did not conduct any post experimental interviews to explicitly test this but

MEMORY JUDGEMENTS

encourage future researchers working on similar designs to assess participant's beliefs about the experimental aims.

We chose to examine believability and reliability judgements separately in two experiments as we felt that asking participants to rate 16 vignettes would have been repetitious and provided more chance for participants to guess the experimental aims. As such, although we found different effects of detail and emotion on believability and reliability, such differences may simply be due to variation across participant groups. We encourage future research investigating constructs of believability and reliability to examine the constructs using within-subjects designs.

Null results found in our experiments, particularly around the manipulation of emotion may in part be due to crime type chosen. As we noted in our introduction, emotion has been found to have a strong effect on credibility, with emotional witnesses often viewed as more credible than those that are non-emotional (Ask & Landstrom, 2010). However, much of the research surrounding this effect has been conducted in crimes of a sexual nature. We wanted to understand if the effect would be replicated in non-sexual crimes and hence chose a crime that we believe is unlikely to be sexually motivated. Future research could compare assessments of believability and reliability across different crime types to see if the effects of emotion and detail produce similar results.

Finally, we encourage future research investigating judgements of the memories of others to include manipulations of account length. As seen in Experiment 1, word count was the only predictor to have a significant effect on assessments of believability. Systematic experimental work using word count as an independent variable may produce some interesting and important findings.

Conclusions

MEMORY JUDGEMENTS

Our findings have important forensic repercussions, which are relevant to cases when individuals lacking expert knowledge about memory are required to judge evidence based upon memory. Our results also suggest that believability (truthfulness) and reliability (accuracy of memory) are distinct constructs. We propose that future work assessing judgements about other people's memories should investigate believability and reliability since both are likely to play an important role throughout the legal process. We have shown how reliability judgements of others' memories are malleable and can be influenced by the inclusion of a few minor details. In support of, and in extending the previous literature, we have shown that assessing memory is a complex and multi-faceted task. This task is partly dependent on erroneous beliefs about how memory operates. Work such as ours, then, is of paramount importance in legal settings where scientific memory research is often pervasively denied.

References

- Amodio, D. M., Jost, J. T., Master, S. L., & Yee, C. M. (2007). Neurocognitive correlates of liberalism and conservatism. *Nature Neuroscience*, *10*(10), 1246-1247. DOI: 10.1038/nn1979
- Arana, J. M., Meilan, J. J. G. and Perez, E. (2008), The effect of personality variables in the prediction of the execution of different prospective memory tasks in the laboratory. *Scandinavian Journal of Psychology*, *49*, 403–411. DOI:10.1111/j.1467-9450.2008.00671.x
- Areh, I., & Umek, P. (2004). Personal characteristics and validity of eyewitness testimony. In G. Meško, M. Pagon, & B. Dobovšek (Eds.), *Policing in Central and Eastern Europe: Dilemmas of contemporary criminal justice* (pp. 355-359). Faculty of Criminal Justice and Security, University of Maribor.
- Ask, K., & Landström, S. (2010). Why emotions matter: Expectancy violation and affective response mediate the emotional victim effect. *Law and Human Behavior*, *34*(5), 392-401. DOI: 10.1007/s10979-009-9208-6.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9 <http://CRAN.R-project.org/package=lme4>.
- Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: the power of (a few) minor details. *Journal of Personality and Social Psychology*, *56*(5), 669-679. DOI: 10.1037/0022-3514.56.5.669.
- Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, *20*(1), 115-129. DOI: 10.1002/acp.1171

MEMORY JUDGEMENTS

- Bollingmo, G. C., Wessel, E. O., Eilertsen, D. E., & Magnussen, S. (2008). Credibility of the emotional witness: A study of ratings by police investigators. *Psychology, Crime & Law, 14*(1), 29-40. DOI: 10.1080/10683160701368412
- Bollingmo, G., Wessel, E., Sandvold, Y., Eilertsen, D. E., & Magnussen, S. (2009). The effect of biased and non-biased information on judgments of witness credibility. *Psychology, Crime & Law, 15*(1), 61-71. DOI: 10.1080/10683160802131107
- Bornstein, B. H. (2017). *Popular Myths about Memory: Media Representations Versus Scientific Evidence*. Lanham, MD: Lexington Books.
- Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53*(4), 594-628. DOI: 10.1016/j.jml.2005.08.005
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*(2), 261-288. DOI: 10.1037/0033-295X.107.2.261
- Conway, M. A., Justice, L. V., & Morrison, C. M. (2014). Beliefs about autobiographical memory and why they matter. *Psychologist, 27*(7), 502-505.
- Cramer, R. J., Brodsky, S. L., & DeCoster, J. (2009). Expert witness confidence and juror personality: Their impact on credibility and persuasion in the courtroom. *Journal of the American Academy of Psychiatry and the Law Online, 37*(1), 63-74.
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*(4), 346-357.
- Ernberg, E., & Landström, S. (2016). To prosecute or not to prosecute: Law students' judicial decisions in CSA cases. *Scandinavian Journal of Psychology, 57*(1), 30-35. DOI: 10.1111/sjop.12261

MEMORY JUDGEMENTS

- Ernberg, E., Tidefors, I., & Landström, S. (2016). Prosecutors' reflections on sexually abused preschoolers and their ability to stand trial. *Child Abuse & Neglect*, *57*, 21-29. DOI: 10.1016/j.chiabu.2016.06.001
- Flehmig, H. C., Steinborn, M., Langner, R., & Westhoff, K. (2007). Neuroticism and the mental noise hypothesis: Relationships to lapses of attention and slips of action in everyday life. *Psychology Science*, *49*(4), 343-360.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26-42. DOI: 10.1037/1040-3590.4.1.26
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528. DOI: 10.1016/S0092-6566(03)00046-1
- Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology*, *9*(1), 103-119. DOI: 10.1348/135532504322776889
- Hirsh, J. B., DeYoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin*, *36*(5), 655-664. DOI: 10.1177/0146167210366854
- Heuer, F., & Reisberg, D. (1992). Emotion, arousal, and memory for detail. In S.-Å. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 151-180). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Hyman, I. E., & James Billings Jr, F. (1998). Individual differences and the creation of false childhood memories. *Memory*, *6*(1), 1-20. DOI: 10.1080/741941598
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67.

MEMORY JUDGEMENTS

- Johnson, M. K., Bush, J. G., & Mitchell, K. J. (1998). *Interpersonal reality monitoring: Judging the sources of other people's memories*. *Social Cognition*, 16(2), 199-224. doi:<http://dx.doi.org/10.1521/soco.1998.16.2.199>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological Assessment*, 22(2), 420-432. DOI: 10.1037/a0019265
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political Conservatism as Motivated Social Cognition. *Psychological Bulletin*, 129(3), 339-375. DOI: 10.1037/0033-2909.129.3.339
- Kanai, R., Feilden, T., Firth, C., & Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, 21(8), 677-680. DOI: 10.1016/j.cub.2011.03.017
- Katz, C., Paddon, M. J., & Barnett, Z. (2016). Emotional language used by victims of alleged sexual abuse during forensic investigation. *Journal of Child Sexual Abuse*, 25(3), 243-261. DOI: 10.1080/10538712.2016.1137666
- Kaufmann, G., Drevland, G. C., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology*, 17(1), 21-34. DOI: 10.1002/acp.842
- Landström, S., Ask, K., Sommar, C., & Willén, R. (2015). Children's testimony and the emotional victim effect. *Legal and Criminological Psychology*, 20(2), 365-383. DOI: 10.1111/lcrp.12036
- Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology*, 63(2), 181-197. DOI: 10.1037/0022-3514.63.2.181

MEMORY JUDGEMENTS

Levesque R. J. R. (2006). *The psychology and law of criminal justice processes*. Hauppauge, New York: Nova Science Publishers.

Magnussen, S., Andersson, J., Cornoldi, C., De Beni, R., Endestad, T., Goodman, G. S. & Nilsson, L. G. (2006). What people believe about memory. *Memory*, *14*(5), 595-613.
DOI: 10.1080/09658210600646716

McGuire, K. & London, K. (2017). Common beliefs about child sexual abuse and disclosure: A college sample. *Journal of Child Sexual Abuse*, *26*, 175-194. DOI:
10.1080/10538712.2017.1281368

Niedźwieńska, A., Neckar, J., & Baran, B. (2007). Development and validation of the implicit memory theory scale. *European Journal of Psychological Assessment*, *23*(3), 185-192. DOI: 10.1027/1015-5759.23.3.185

Ost, J., Easton, S., Hope, L., French, C. C., & Wright, D. B. (2017). Latent variables underlying the memory beliefs of Chartered Clinical Psychologists, Hypnotherapists and undergraduate students. *Memory*, *25*(1), 57-68. DOI:
10.1080/09658211.2015.1125927

Patihis, L., Ho, L. Y., Tingen, I. W., Lilienfeld, S. O., & Loftus, E. F. (2014). Are the “memory wars” over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science*, *25*(2), 519-530.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556-563. DOI: 10.1016/S0092-6566(02)00505-6

Penrod, S. D., & Cutler, B. (1999). Preventing mistaken convictions in eyewitness identification trials. In *Psychology and Law* (pp. 89-118). Springer: Boston, US.

MEMORY JUDGEMENTS

Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). Negotiating false memories: Interviewer and rememberer characteristics relate to memory

distortion. *Psychological Science*, *11*(6), 507-510. DOI: 10.1111/1467-9280.00297

Reyes, R. M., Thompson, W. C., & Bower, G. H. (1980). Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social*

Psychology, *39*(1), 2-12. DOI: 10.1037/0022-3514.39.1.2

Rosenthal, J. R. (2002). Suggestibility, reliability, and the legal process. *Developmental Review*, *22*(3), 334-369.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA

URL <http://www.rstudio.com/>.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182-203. DOI: 10.1037/0003-

066X.54.3.182

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the US population. *PloS one*, *6*(8), e22757. DOI:

10.1371/journal.pone.0022757

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P.-A. Granhag & L. Strömwall (Eds.), *The detection of deception in*

forensic contexts (pp. 229-250). DOI: 10.1017/CBO9780511490071.010

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, *18*(1), 46-50.

DOI: 10.1111/j.1467-9280.2007.01847.x

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley & Sons.

MEMORY JUDGEMENTS

- Wells, G. L., & Leippe, M. R. (1981). How do triers of fact infer the accuracy of eyewitness identifications? Using memory for peripheral detail can be misleading. *Journal of Applied Psychology, 66*(6), 682-687. DOI: 10.1037/0021-9010.66.6.682
- Wessel, E. M., Bollingmo, G. C., Sønsteby, C., Nielsen, L. M., Eliersten, D. E., & Magnussen, S. (2012). The emotional witness effect: Story content, emotional valence and credibility of a male suspect. *Psychology, Crime & Law, 18*(5), 417-430. DOI: 10.1080/1068316X.2010.502119
- Wessel, E. M., Eilertsen, D. E., Langnes, E., Magnussen, S., & Melinder, A. (2016). Disclosure of child sexual abuse: expressed emotions and credibility judgments of a child mock victim. *Psychology, Crime & Law, 22*(4), 331-343. DOI: 10.1080/1068316X.2015.1109087
- Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18*(4), 427-443. DOI: 10.1002/acp.993.
- Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger III, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*(6), 515.
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*(1), 10-65. DOI: 10.1177/1529100616686966
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... & Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences, 48*(8), 889-894. DOI: 10.1016/j.paid.2010.02.016

MEMORY JUDGEMENTS

Appendix 1

Memory Vignettes

1 Low detail, non-emotional

It was June 27th. I was in a coffee shop when someone ran up behind me and pushed me to the ground. I realised that my lap top had been stolen. I don't remember much, just that he was dressed in darkish clothes. He might have shouted something at me. I felt quite unafraid.

1a Low detail, non-emotional

It was March, the 22nd. I was on the way to work he was standing very close then suddenly pulled my jacket so hard it ripped. I noticed then he had taken my car keys from my pocket. I don't remember it very clearly, I think he had some kind of tattoo. I think he muttered something. I wasn't very scared.

2 Low detail, emotional

The day was 14 October. I was in a restaurant and he walked up to the table and jammed it towards me He had taken my bank card. I hardly remember anything about what happened, I just remember that he probably had fairish hair. He may have spoken to me. I felt absolutely terrified.

2a Low detail, emotional

The 8th of May, I was on a train. He turned around and bumped into me so hard I fell over. I found out after that my passport had been taken. My memory for the event is very hazy. I didn't notice if he had any piercings. He probably said something to me. I was completely petrified.

3 High detail, non-emotional

It happened in December, the 12th. I was standing up on a bus and he was walking towards me to get off and aggressively barged straight into me. As he was leaving I realised that he had stolen my ipad I remember everything, I remember that he was wearing a dark brown jacket and that he shouted "shut up". I felt relatively calm.

3a High detail, non-emotional

It was the 7th of September, and I was in a shopping centre. The guy appeared in front of me and grabbed me by the arm. That must have been when he stole my watch. I remember all of what happened. He had a black tribal tattoo on his forearm. He spoke, and said "don't move". I wasn't that frightened.

4 High detail, emotional

On the 3rd of February, I was going for a walk. He came from nowhere, shoved me over and stole my phone. My memory of the event is very vivid. He had wiry blonde hair and muttered "give me your phone now". I felt utter panic.

4a High detail, emotional

The event happened on the 12th of July whilst I was in a queue for the checkout. He was standing very close behind me and purposely tripped me over. I went to pay later and realised that all my money had been stolen. I remember it so clearly. He had a silver stud in his left ear. He said I should keep quiet. I was so frightened.

MEMORY JUDGEMENTS

	Fixed	Experimental: high	Experimental: low
Date of event	On the 3 rd of February	It was June 27 th	
Current activity	On the way to work	Coffee shop	
Criminal event	He came out of nowhere, shoved me over	He ran up behind me, pushed me to the ground	
Stolen item	Car keys	Laptop	
Recall statement	I remember everything	I don't remember much	
Detail: perpetrator	He was wearing a dark brown jacket	He was dressed in darkish clothes	
Detail: speech	He shouted "shut up"	He might have shouted something at me	
Emotionality statement	I felt utter panic	I felt relatively calm	

Figure 1. Example fixed and experimental vignette elements

MEMORY JUDGEMENTS

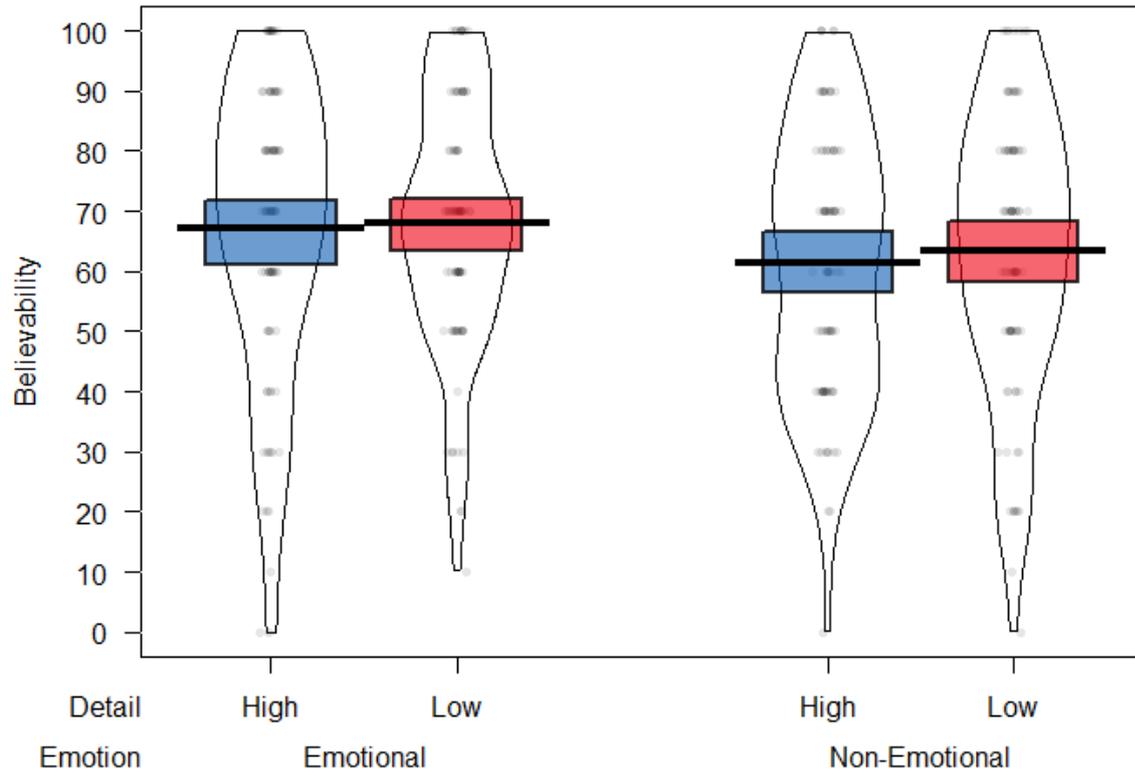


Figure 2. Believability ratings as a function of detail and emotion. Chart elements: points = individual data points, bar = sample mean, band = 95% highest density interval of the population mean, outline = smoothed density.

MEMORY JUDGEMENTS

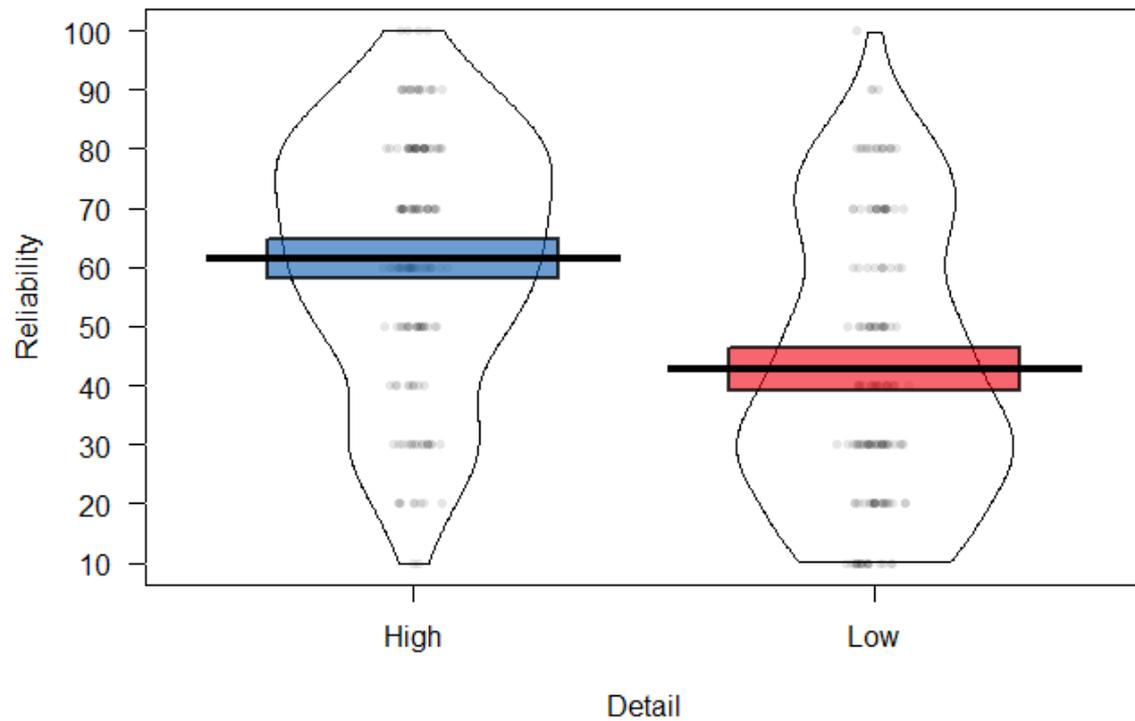


Figure 3. Reliability ratings as a function of detail. Chart elements: points = individual data points, bar = sample mean, band = 95% highest density interval of the population mean, outline = smoothed density.

MEMORY JUDGEMENTS

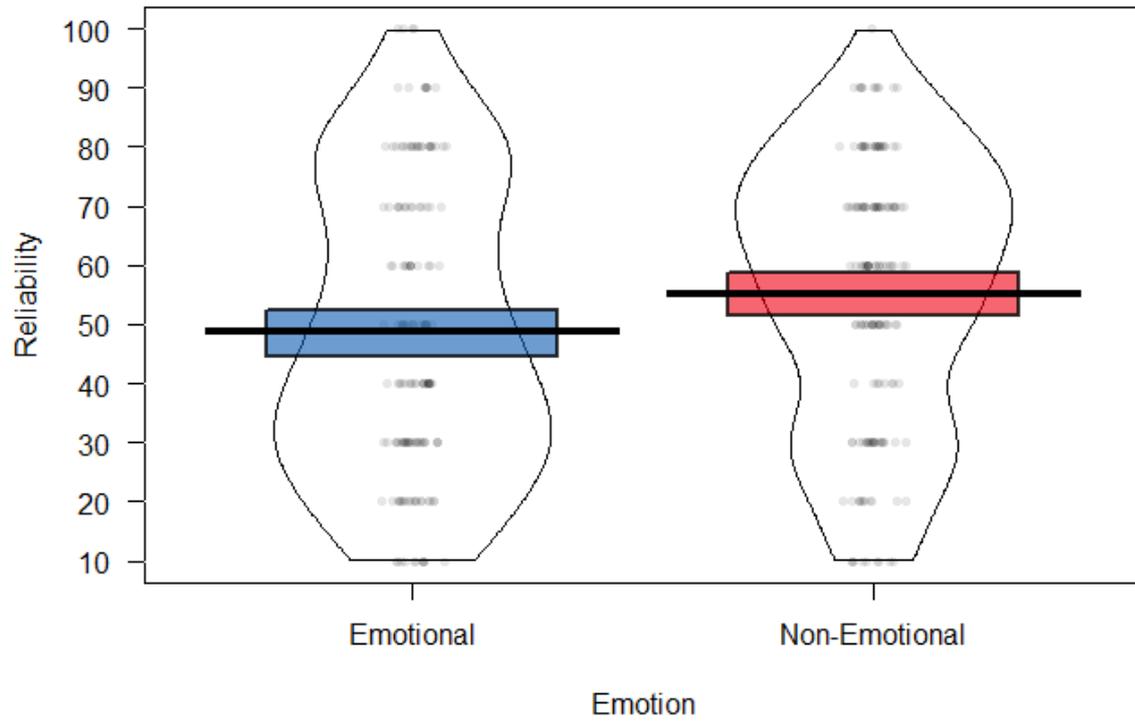


Figure 4. Reliability ratings as a function of emotion. Chart elements: points = individual data points, bar = sample mean, band = 95% highest density interval of the population mean, outline = smoothed density.

MEMORY JUDGEMENTS

Table 1. Means (and standard deviations) of believability and self-reported confidence for each vignette

Vignette	Detail	Emotion	Believability	Confidence
1	Low	Non-emotional	63.02 (24.35)	60.47 (24.68)
1a	Low	Non-emotional	63.49 (23.39)	58.14 (24.13)
2	Low	Emotional	66.59 (21.78)	60.23 (24.35)
2a	Low	Emotional	69.55 (20.34)	58.18 (24.04)
3	High	Non-emotional	63.18 (20.66)	58.64 (23.19)
3a	High	Non-emotional	59.53 (24.1)	60.7 (24.53)
4	High	Emotional	77.44 (17.2)	61.63 (26.09)
4a	High	Emotional	57.05 (27.41)	55.91 (24.81)

MEMORY JUDGEMENTS

Table 2. Means (and standard deviations) of reliability and self-reported confidence for each vignette

Vignette	Detail	Emotion	Reliability	Confidence
1	Low	Non-emotional	48.86 (21.66)	65.43 (15.78)
1a	Low	Non-emotional	48.29 (22.56)	63.71 (16.18)
2	Low	Emotional	40.88 (21.79)	61.18 (18.05)
2a	Low	Emotional	42.57 (24.89)	61.43 (18.65)
3	High	Non-emotional	64.12 (20.91)	65.88 (19.56)
3a	High	Non-emotional	64.29 (19.6)	65.14 (16.34)
4	High	Emotional	60.29 (21.39)	62.65 (19.43)
4a	High	Emotional	58.57 (24.75)	61.43 (20.02)