The Role of Recollection in the Confidence-Accuracy Relationship for Negative Decisions

Sarah Fletcher (nee Timmins) BBSc (Hons.)

School of Psychology

Faculty of Social Sciences

A thesis submitted to Flinders University for the degree of Doctor of Philosophy

December, 2014

TABLE OF CONTENTS

SUMMARY

This research used a series of word recognition paradigms to investigate the relationship between post decision confidence and accuracy for negative recognition decisions (identifying a stimulus as unseen). The experiments focussed on a specific factor that may impact on this relationship: the extent to which individuals used recollected evidence, as described by dual-process theories of recognition. This is important for understanding in which situations and for which types of decisions confidence may be used to predict accuracy. Negative recognition decisions have previously been found to have weaker confidence-accuracy relationships than positive recognition decisions (recognising that a stimulus has been seen) in certain tasks. The aim of this research was to discover whether this positive-negative difference also occurs in word recognition tasks and to investigate the role of use of recollection in this relationship.

The experiments used word recognition paradigms including item recognition, plurality discrimination and an opposition procedure involving read and heard words to investigate the relationship between confidence and accuracy for negative decisions when different amounts of recollected evidence were available. This is important as an understanding of the relationship between the type of evidence used in recognition decisions and the capacity of confidence to predict accuracy allows prediction of when confidence may be used to indicate accuracy and when it may not. This is influential in many fields were human recognition decisions have weighty consequences, and therefore estimating the likely accuracy of a decision maker is desirable.

In the first set of experiments I compared tasks expected to vary in the use of recollection due to the availability of recollected evidence in the task and the degree to which participants were expected to view it as useful. Results demonstrated that in the tasks where recollected evidence was expected to be used more, the relationship between confidence and accuracy for

negative decisions was stronger. The second set of experiments manipulated recollection to ensure differences were not due to other discrepancies between the paradigms. These experiments demonstrated that when recollection was impaired, the confidence-accuracy relationship for negative decisions was also impaired. The final experiment attempted to manipulate recollection in the reverse direction by increasing the availability of recollected evidence for some decisions. Results demonstrated that the confidence-accuracy relationship for negative decisions was strengthened and the positive-negative difference was reduced when recollected evidence was made more available.

These results have implications for a) the situations in which confidence may be used as a marker of accuracy and b) how recognition memory testing situations might best produce the strongest possible confidence-accuracy relationship for negative decisions. They also demonstrate that the positive-negative difference generalises across recognition tasks and is therefore likely to be based on underlying differences in basic cognitive processes. I suggest that neglect of recollected evidence in recognition tasks may be an important cause of poor confidence-accuracy relationships for negative compared with positive decisions, and therefore future research should aim to investigate methods of increasing recollection in important recognition tasks, and the impact these manipulations have on the confidence-accuracy relationship.

DECLARATION

I certify that this thesis does not contain any material that has been accepted for the award of any other degree or diploma; and that to the best of my knowledge and belief does not contain any material previously published or written by another person except where due reference is made in the text of the thesis or notes.

S. Fletcher BBSc (Hons.)

ACKNOWLEDGEMENTS

This research would not have been possible without the support of many influential advisors, friends, family and colleagues.

Thankyou Nathan for your wisdom and honest feedback and for pushing me to think to depths I had almost forgotten how to at the end, and to Neil, who provides our lab and its wonderful culture and work ethic, and Paul, Matt, Jason and Michelle for believing in my research abilities and always passing on ideas or links that were helpful.

To all my lab mates and officemates: especially Nicole for being my lab mentor and showing me the ropes; Stacey, for understanding my temperament and holding my hand when I had to make an inquiry at an office or visit the school finance officer, as well as being company and conversation in the office whenever I needed it, and frequently providing delicious baked research fuel; Nicky, for being my second and third year office sounding board and for your ideas and inspiration which helped me to remain passionate about the research we do; and Rosie for being my non-memory office buddy and sometimes much needed distraction, thankyou. Thankyou also to Tomoko for enduring my last minute ‘how to hand up’ questions and for along with Caitlin, Nicole and Rachel being my first overseas conference escort.

To my husband Lloyd, for being the bearer of tea and hugs, love and a fun and refreshing place to call home, and breaking up my PhD with the joy of our wedding and the best dates and presents ever, while still helping me with proofreading, maths, IT, and logic, and to both my own parents, for maintaining the music I dance my life to, and those I acquired through marriage, for your support in all my endeavours and interest in my work. To my late Grandma for assisting me through undergraduate studies and teaching me the value of a cup of tea, which keeps me going to this day through all challenges, because tea’s what it’s all about. And to my fur children, Taranis and Lucky, for your unconditional love.

To my adult Irish dance class, for your friendship, laughter, fitness inspiration, wonderful achievements and constant reassurance of my competence in fields outside study, and for helping me to achieve other dreams at the same time as this one: Michelle, Elise, Kate, Lauren, Nicole, Maddy, Clare, Mum, Emilia, Louis, and Miranda, and to Deirdre, for making work with them possible. You have all provided me with a source of health, social support, and the most wonderful friendship throughout the duration of this work.

And to my outside of PhD friends, especially my bridesmaids, Jess, Emily and Mel, who have been a continuous support team and source of inspiration, and the wonderful storyteller Ashton, who transported me to another place once a week for all of these past four years, and inspired my creative spirit.

Thankyou all of you from the bottom of my heart.

CHAPTER 1 — INTRODUCTION

Decisions based on memory form a crucial part of our everyday lives. A common way of understanding the varied memory tasks that individuals complete is to broadly classify them into two types: recall memory and recognition memory. Recall is used when an individual is required to produce and report information from memory, for example, answering free recall questions such as ‘what did you have for breakfast?’ Recognition is used when deciding if items have been seen before and/or whether they are the particular item sought. Recognition memory was the focus of this research.

A recognition decision is made whenever a person has to determine whether they have seen an item before (Mandler, 1980; D. A. Norman & Wickelgren, 1969). Recognition decisions determine whether an individual can identify their luggage when leaving the airport, identify where they know a familiar face from, or find a friend’s car in a busy carpark. Although the examples above are relatively mundane ones, these types of decisions can sometimes have far reaching consequences. Some examples of specific contexts in which recognition judgments have critical outcomes include doctors examining medical tests for symptoms (Boutis, Pecaric, Seeto, & Pusic, 2010); eyewitnesses identifying criminals from a police lineup (Wells, 1993); or security staff checking baggage (Hardweier, Hofer, & Schwaninger, 2005; McCarley, 2009) or watching for wanted criminals among a crowd. In each case, an individual must decide whether an item is present which matches a case seen in the past. If an incorrect decision is made in any of these examples consequences may be severe. For example, if the doctor decides that symptoms of serious disease are absent when the test does show symptoms, a patient may suffer.

Because of the impact that some recognition decisions like these can have, as well as the theoretical interest of understanding recognition as a cognitive process, it is useful to investigate variables that predict the accuracy of recognition decisions. For this reason one branch of memory research has focussed on finding predictor variables that can help estimate the likelihood that a given recognition decision is correct (Sauerland & Sporer, 2009; Sporer, 1992; Sporer, Penrod, Read, & Cutler, 1995). Two characteristics of recognition decisions that have sometimes been found to predict accuracy are post-decision confidence in the decision's accuracy, (Balakrishnan & Ratcliff, 1996; Brewer & Wells, 2006; Lindsay, Read, & Sharma, 1998; Sauerland & Sporer, 2009) and response latency, the time taken by the individual to make their recognition decision (Brewer & Weber, 2008; Sauerland & Sporer, 2009). However these variables do not predict the accuracy of recognition decisions in the same way in all situations.

Research in the field of eyewitness identification and face recognition has found that the ability of these variables to predict accuracy depends on the type of decision made (e.g., Brewer & Weber, 2008; Weber & Brewer, 2006). In a forced choice recognition test (where individuals are required to answer yes or no to indicate whether an item has been seen before) two kinds of decision can be made about a presented item. First, a positive decision, that the item has been studied or seen before. Second, a negative decision, that the item has not been seen or studied. Positive decisions that are more confident and more rapid are more likely to be correct. However for negative decisions, confidence and response latency are consistently more weakly, and sometimes simply not related, to accuracy (e.g., Brewer & Wells, 2006; Sauerland & Sporer, 2009; Sporer, 1992; Sporer et al., 1995; Weber & Brewer, 2006). An explanation of this variation in the confidence-accuracy relationship was the main focus of this thesis[1]. Specifically an explanation based on dual-process theories of recognition, which will be explained in more detail below, was tested.

Dual-process theories propose that two kinds of evidence can underlie a recognition decision. The first, familiarity, is a general match of an item to all (relevant) material in memory. The second, recollection, has been conceived of differently in different theories, but generally is considered to involve the use of specific details retrieved from memory (Cary & Reder, 2003; Jacoby, 1991; A. P. Yonelinas, 1994, 1999). Here I will use the definition (following, e.g., Malmberg, 2008), that it involves actively recalling item-specific information from memory that matches or mismatches the item being viewed. I suggest and test the idea that the weaker ability of confidence to predict the accuracy of negative recognition decisions compared with positive is due to neglect of recollected evidence. However, before going into detail about the proposed explanation I will present the literature that has established the positive-negative difference as a consistent phenomenon and review the available theories of recognition memory.

Evidence for a Positive-Negative Difference

A number of studies in the areas of eyewitness identification and face recognition have established a consistent difference in the confidence-accuracy relationship between positive and negative decisions. The first study to note the positive-negative difference was conducted by Sporer (1992). Participants in Sporer's study witnessed an intruder attempting to take away the projector during a laboratory exercise. Participants were later called back to the laboratory and viewed video lineups in which they were asked to attempt to identify the

[1] Response latency will also be measured and results concerning it presented for potential future use in construction of a more complex model, but it will not be the main focus of investigation.

intruder. For those participants who made a positive decision, indicating someone from the lineup was the intruder, post-decision confidence was moderately strongly correlated (r = 0.58) with accuracy. However, for those participants who made a negative decision, indicating that the intruder was not present in the lineup, post-identification confidence and accuracy were not significantly correlated, and the correlation was markedly smaller (r = 0.08).

Sporer et.al. (1995) subsequently conducted a meta-analysis of previous eyewitness identification studies that had concluded that the relationship between confidence and accuracy was weak and, consequently, recommended that confidence not be viewed as a reliable indicator of accuracy. By introducing decision type (whether the participant made a positive or negative decision) as a moderating variable, Sporer et.al. discovered that the confidence-accuracy relationship was significantly stronger when participants made a positive decision (r = .37) than when they made a negative decision (r = .12).

Brewer and Wells (2006) replicated the positive-negative difference for the confidence-accuracy relationship using an eyewitness identification paradigm. Participants viewed a video in which a credit card was taken from a restaurant, and were then asked to identify the thief from a lineup consisting of 8 colour photographs, or to click on the ‘not present’ button if they believed the thief was not in the lineup. Following their decision they were asked to indicate their confidence from 0-100%. The confidence-accuracy relationship was examined using two approaches which indicate different properties of the confidence-accuracy relationship. The first, calibration, indicates the extent to which individuals’ confidence ratings are realistic. The calibration approach compares the proportion correct with the respective confidence rating for a category, such that perfect calibration occurs when the confidence value matches the probability that a decision is correct. That is, perfect calibration occurs if 50% of judgments made with 50% confidence are correct, and so on for each confidence category. The calibration statistic (*C*) ranges from 0 (perfect calibration) to 1 (no calibration), Resolution, quantified by the Normalised Resolution Index (*NRI*, or Adjusted Normalised Resolution Index, *ANRI*) indicates whether confidence can be used to discriminate accurate from inaccurate decisions on a scale of 0 (no discrimination) to 1 (perfect discrimination). Both measures are covered in more detail in chapter 2. In Brewer and Wells' study, resolution for negative decisions was close to zero (.002 for one stimulus and .043 for the other), while for positive decisions resolution was above .10 for both stimuli (.107, .132). Although confidence-accuracy calibration was similar for positive and negative decisions, calibration is meaningless when resolution is zero, as it may indicate for example that 50% of an individuals' decisions were correct and they only ever used a confidence rating of 50%. Hence Brewer and Wells (2006) found that confidence could be used to discriminate correct from incorrect positive decisions, but not correct from incorrect negative decisions.

Sauerland and Sporer (2009) also replicated the positive-negative difference in confidence in a lineup field study. Passers-by were asked for directions by a confederate and were later asked to identify that person from a photo lineup. For positive decisions, confidence demonstrated a moderate positive correlation with accuracy (r = .39), good calibration between confidence and accuracy (C = .026), and high resolution (*NRI* = 0.174). This shows that confidence ratings were realistic (i.e., reflected the probability that decisions were accurate) and confidence could be used to distinguish between accurate and inaccurate decisions. However for negative decisions, the confidence-accuracy correlation was not significant (r = .09), confidence-accuracy calibration was poor (C = .059), and resolution was low (*NRI* = .030), indicating limited realism of confidence judgments and little ability to discriminate accurate from inaccurate decisions on the basis of confidence. Sauer, Brewer, Zweck and Weber (2010) found similar results for an eyewitness identification paradigm with positive decisions displaying higher resolution than negative decisions.

Many face recognition studies have also found evidence of a positive-negative difference, although the impairment in confidence-accuracy resolution for negative decisions is often not as severe. Weber and Brewer (2003) asked participants to study photographs of faces and later make recognition judgements about whether faces had been seen in the study phase. Results showed that regardless of other conditions, confidence-accuracy calibration was superior when participants made positive decisions (indicating they had seen the face before) than when they made negative decisions (indicating they had not seen the face before). In a subsequent studies (Weber & Brewer, 2004, 2006) resolution was also assessed and demonstrated consistently superior resolution for positive decisions despite similar calibration. This result is consistent with Brewer and Wells' data and suggests that the positive-negative difference occurs at the level of resolution, which can in turn sometimes produce poor calibration.

Overall, in every case where the choice participants made (positive or negative decision) has been considered in analysing the results, both face recognition and eyewitness identification studies have demonstrated that participants' level of confidence in their decision is related to their accuracy when they make a positive decision. However, where participants state that a face presented to them has not been previously seen, or none of the faces in a lineup belong to the culprit, their confidence in this decision is minimally or sometimes not at all indicative of their accuracy. In particular, poor confidence-accuracy resolution is shown, such that confidence does not discriminate correct from incorrect negative decisions. This suggests a difference in the underlying cognitive mechanisms, or the evidence input into those memory processes, between positive and negative recognition judgements and confidence ratings, which this research aimed to expose.

Response Latency and Accuracy

In addition to the positive-negative difference occurring in the confidence-accuracy relationship, many of the studies discussed in the previous section also demonstrated that the positive-negative difference occurs in the response latency-accuracy relationship. For positive decisions, response latency and accuracy display a negative relationship, while for negative decisions, this relationship is weaker (Lichtenstein & Fischhoff, 1977; Sporer, 1992; Weber, Brewer, Wells, Semmler, & Keast, 2004) The fact that the difference between positive and negative decisions is common to the response latency-accuracy and confidence-accuracy relationships suggests that it is caused by either a difference in the way positive and negative recognition decisions are made, or a difference in the type or quality of information used to make positive and negative decisions. Confidence and response latency are often related, and this is thought to be because they are caused by the same underlying information. Given that the confidence- accuracy and response latency-accuracy relationships vary in the same way between positive and negative decisions, therefore, a difference in the underlying information producing both confidence and response latency is likely to exist between positive and negative decisions. This section briefly summarises the evidence that the positive-negative difference also occurs in the response latency-accuracy relationship. However, the confidence-accuracy relationship will be the primary focus of the thesis.

Sporer (1992) noted that decision type also moderated the response-latency-accuracy relationship in eyewitness recognition. Positive identifications from the video lineup displayed a negative correlation between response latency and accuracy (r = -.43). However, for negative decisions, the correlation between latency of responses to the lineup and identification accuracy was smaller (r = -.20) and was not significant. In a later study by Sporer in (1993), response-latency was negatively correlated with accuracy when participants indicated that someone in the lineup was the offender (that is, they made a positive decision, *r* = - .36). However, when they indicated that the offender was not present, the correlation between response latency and accuracy was again smaller and non-significant (*r* =.17). A difference in the response latency-accuracy relationship for positive and negative decisions, with positive decisions showing a stronger relationship, was also found by Kneller, Memon, and Stevenage (2001) who showed that positive identifications from sequential lineups (*M* = 3.82, *SD* = 2.87) were made significantly faster than inaccurate foil selections (*M* = 16.45, *SD* = 13.17). However, accurate lineup rejections (*M* = 3.87, SD = 2.13) were not made significantly faster than inaccurate lineup rejections (*M* = 3.95, *SD* = 1.78).

A difference in the response latency-accuracy relationship has also been shown for face recognition. Weber and Brewer (2006) directly compared the response latency-accuracy relationship for positive and negative decisions, and found that although response latency was negatively related to accuracy for all decision types, the relationship was significantly weaker for negative decisions than for positive decisions. Indirect evidence showing that positive decisions display stronger response latency-accuracy relationships than negative decisions has also been shown by Sauerland and Sporer (2009) and by a meta-analysis of four separate eyewitness identification studies by Weber et al. (2004).

Response latency, like confidence, is predicted by many recognition theories (e.g., Gillund & Shiffrin, 1984; Ratcliff, 1978) to be related to the accuracy of all recognition decisions. Early studies looking into the relationship between response latency and accuracy in eyewitness identification found a strong negative relationship; that is, faster decisions were indeed more likely to be accurate (Sauerland & Sporer, 2009; Slotnick, 2010; Sporer, 1992, 1993). Similarly, evidence suggests that strong memories are more rapidly retrieved, (Brewer, Weber, Wootton, & Lindsay, 2012; Gillund & Shiffrin, 1984) suggesting that decisions made based on good memory, which are more likely to be accurate, should also be rapid. Many early theories of decision making suggest that rapid decisions are more likely to be correct, as slower decisions only occur when the individual making the decision has conflicting evidence in mind which does not clearly favour one of the response alternatives (e.g., Baranski & Petrusic, 1994; Festinger, 1943; Gillund & Shiffrin, 1984; Heathcote, Bora, & Freeman, 2010; Jacoby, Jones, & Dolan, 1998), or when evidence is not strong enough in favour of either response.

Taken together these results suggest that an underlying difference in the process leading to positive and negative decisions influences the degree to which both confidence and the latency of decisions are related to their accuracy. Therefore, although I did not aim to make specific predictions about response latency for the experiments in this thesis, I expected manipulations which altered the confidence-accuracy relationship to also affect the response latency-accuracy relationship. Specific predictions about response latency would require use of a detailed model of the time course of recognition decisions, which is beyond the scope of this thesis.[2] However, the response latency-accuracy relationships were examined in order for the data to be available for furthering development of such models in the future, and also to test the basic prediction that the confidence-accuracy and response latency-accuracy relationships would follow similar patterns. An understanding of the response latency-accuracy relationship for negative decisions is desirable as many recognition memory theories support the idea that fast decisions are more likely to be correct (e.g., Baranski & Petrusic, 1998) and, therefore, the fact that negative decisions do not always follow this prediction has implications for the development of theories of recognition.

[2] As the focus of this thesis was on testing a broad conceptual level explanation of the positive-negative difference, selecting and testing one single theory was not desirable. Several of the models available for predicting response latency do not distinguish between recollection and familiarity and were therefore not appropriate to the explanation being considered (e.g., Clark & Gronlund, 1996; Diller, Nobel, & Shiffrin, 2001; Gillund & Shiffrin, 1984), and those that do include both types of memorial evidence (e.g., Juola, Fischler, Wood, & Atkinson, 1971) are not yet well tested or do not attempt to explain response times (e.g., Reder, Angstadt, Cary, Erikson, & Ayers, 2000)

Potential Explanations for the Positive-Negative Difference in Confidence: Explanations Unique to the Eyewitness Task

This section will summarise proposed explanations of the positive-negative difference in confidence, none of which fully explain the phenomenon. These explanations primarily attempt to explain the phenomenon as an outcome of specific features of the eyewitness identification task, as described below.

Different evidence bases. Weber and Brewer (2006) suggested that a possible reason for the differing confidence-accuracy relationship for positive and negative eyewitness identification decisions was the different evidence bases on which the decisions are made. Since a lineup involves more than one stimulus face, an individual must decide that none of a series of faces match their memory for the target, in order to make a negative decision. On the other hand positive decisions, and confidence judgements about them, can be based on the degree of match of a single stimulus item (the best matching face in the lineup) to a single item in memory. In addition, Weber and Brewer suggested that this could apply for face recognition because individuals normally are presented with a series of faces at study, but at test are only given a single face stimulus and asked to decide if it was a) one of the studied faces, or b) did not match any of the studied faces. If the face is one of the studied faces, it can be compared one-to-one with the best matching face in memory and a decision, and confidence rating, can be made based on the match between memory and the test face. However, in a negative decision, when the test face does not match any of the studied faces, confidence cannot be based on the mismatch with a particular face studied, as the test face mismatches all the faces studied. Hence confidence judgements for negative decisions could be based on various different comparisons, rather than a one-to-one comparison of features as used for positive decisions.

Weber and Brewer (2006) investigated this possibility using a task in which all decisions should have been based on one-to-one comparisons; an associative recognition task in which faces were paired with specific cues. Participants were shown a series of pictures of faces each presented with a different name, background colour or occupation. At test they were presented with these cues (the name, occupation or background colour) along with a single face, and asked to determine whether the face was the one presented with that cue during study. Hence participants should always have been assessing the match of the test stimulus to their memory of the face presented with a specific cue. A control condition involving item recognition was also included in which participants simply indicated whether or not single test faces were from the list presented at study. In both conditions participants studied multiple faces and then made decisions at test about individual faces. However, when the cues were added, only one of the faces at study was relevant to the test; the one presented with a particular cue. Therefore, negative decisions could be expected to be based on mismatch between the test stimulus and the face presented with the same cue during study. After each decision participants rated their confidence on a scale of 50% - 100%. Results showed that in the associative recognition task as well as the item recognition task, confidence-accuracy calibration was superior for positive decisions when compared with negative decisions. Hence the difference in the confidence-accuracy relationship was not explained by the different evidence bases underlying the decisions. However as there was no manipulation check to indicate the extent to which participants really used one-to-one match or mismatch in their decisions, participants may have made decisions in both the item and associative tasks in the same way. This is more likely given that the associative test included new items that the participants had not studied, and thus at least some studied and unstudied items could be distinguished without assessing match or mismatch to a particular item in memory.

Sauerland, Sagana, and Sporer (2012) conducted an experiment with the same purpose as Weber and Brewer (2006) using showups (one person lineups) to ensure that all decisions were based on the comparison of a single stimulus face to a face in memory. Participants viewed short films and were then asked to indicate whether a face shown to them was one seen in the video. Under these conditions, Sauerland et.al. found that confidence and accuracy were well calibrated for participants making negative decisions. This contrasts with earlier studies that found no significant relationship between confidence and accuracy for negative decisions. However, as the confidence-accuracy results for positive decisions were not reported in this study, a positive-negative difference could still have occurred in the strength of the relationship. Sauerland et. al.'s results suggest that in some situations, allowing negative decisions to be based on one to one comparisons may result in a useful degree of calibration between confidence and accuracy. However they do not indicate whether the positive-negative difference may be reduced.

Heterogeneous reasons for negative decisions. Sauerland and Sporer (2009) suggested that the reason for the positive-negative difference could be that while there is only one reason to choose someone from a lineup, there are several reasons to say the offender is not present: a) being certain the person in memory is not in the lineup b) being familiar with someone in the lineup but not being confident enough that they are the culprit to identify them, and c) having no clear memory of the event, and hence not recognising anyone in the lineup. Sauerland and Sporer (2009) hypothesised that when individuals made negative decisions due to certainty the target was not in the lineup, confidence and response latency would both be more strongly related to accuracy than when a negative decision was made for either of the other two reasons, which could both indicate weak memory. Sauerland and Sporer asked individuals who rejected the lineup to indicate which of these reasons led them to their decision. Confidence was positively correlated and well calibrated with accuracy for those participants who made a positive decision, but not for those who rejected the lineup. Of those making a negative decision, participants who said they rejected the lineup because the target was not there were more confident in their decisions, but were not significantly more accurate than those who rejected the lineup for either of the other reasons. Hence no evidence was found for the suggestion that it is the heterogeneous nature of negative decisions, compared with positive decisions, which causes the positive-negative difference.

To summarise the evidence for proposed explanations of the positive-negative difference, tests of theories that explain the difference on the basis of factors relating specifically to the task of choosing from a lineup or recognising a face have had varied and conflicting results. Sauerland et.al.'s (2012) results show that there are situations in which negative decisions may show good confidence-accuracy calibration, and one of these may be when negative decisions can be based on a single comparison between a stimulus in memory and a stimulus about which a decision needs to be made. However, a) Weber and Brewer (2006) did not find evidence for this explanation and b) calibration is not useful without a non-zero level of resolution, which the confidence-accuracy relationship for negative decisions may not always have (Brewer & Wells, 2006). I will suggest an extension of this explanation based on one of the dominant types of general recognition theories, dual-process theories, which are described below.

Theories of Recognition Memory

Theories of recognition memory suggest that confidence and accuracy should be related, and past research generally has shown a positive relationship. In laboratory recognition tasks, memory strength and confidence have generally been found to be positively correlated (Mandler, 1980; Van Zandt, 2000; Wixted & Stretch, 2004). Hence eyewitness and face recognition are unusual in not always finding support for a relationship between confidence and accuracy (Brewer & Weber, 2008). Theories of recognition usually suggest that the recognition decision and confidence rating about that decision are based on the same evidence, and therefore should be related (e.g., Clark, 1997; Murdock & Dufty, 1972). While few theories have been proposed to explain why a positive-negative difference might occur in recognition in general, some attempts have been made to suggest possible differences in the nature of information provided for positive and negative decisions in face recognition and eyewitness lineup tasks specifically, which will be discussed in later sections.

The aim of my research was to use basic theories of the mechanisms underlying recognition to explain the positive-negative difference, so that any explanation would not be unique to the eyewitness task or face recognition but may explain why this phenomenon would occur in any recognition task that has similar requirements. Moreover, understanding the basic memory mechanisms underlying these types of recognition decisions is likely to provide more basic and generalisable ideas about how to improve or more easily evaluate the accuracy of recognition decisions. In this section I summarise some of the types of theories currently used to explain the process of recognition decisions, and what they contribute towards an explanation of the positive-negative difference. I briefly introduce two broad classes of recognition theories; single process and dual process theories, and the elements that each have to offer to an explanation of the positive-negative difference. I then suggest an explanation based on elements of these existing theories.

Most recognition theories describe a role of familiarity or strength associated with an item in making a recognition decision. Some types of theories, known as single process or global matching theories (e.g., Clark & Gronlund, 1996; Gillund & Shiffrin, 1984; Shiffrin & Steyvers, 1997; Van Zandt, 2000), consider this to be the only type of evidence used in recognition decisions. When familiarity is used to make a recognition decision, the familiarity or strength associated with a test item is compared to a criterion. If familiarity exceeds the criterion, a positive decision is made. If it does not, a negative decision is made (Gillund & Shiffrin, 1984; Hintzman, 1988; Shiffrin & Steyvers, 1997). Alternatively two criteria are set; matches exceeding the higher criterion produce positive responses, matches not exceeding the lower criterion produce negative responses (e.g., TODAM, Murdock, 1982), and items with familiarity falling in between are given a guess or 'don't know' response. This idea is based on Signal Detection Theory (Murdock & Dufty, 1972), and forms part of most theories of recognition. Confidence is then assumed to stem from the distance familiarity falls from the criterion, with greater confidence in positive decisions reflecting more familiarity associated with the item, and greater confidence in negative decisions reflecting less familiarity with the item (e.g., Clark, 1997; Murdock & Dufty, 1972).

The idea that individuals use the distance familiarity falls from a decision criterion to make confidence ratings suggests a possible explanation for why positive and negative decisions might differ in their confidence-accuracy relationships. Consider the description of confidence as representing distance familiarity falls from the decision criterion. The degree of familiarity with a test item represents the amount of evidence that the item has been seen. When a high confidence positive decision is made, this is based on high familiarity: A large amount of evidence for a positive decision. However, when familiarity is low, and therefore there is not enough evidence for a positive decision, a negative decision may be made based on a smaller amount of evidence. In the first case the greater the amount of evidence for the decision type, the more confident the decision. However, in the second case, there is no direct evidence for a negative decision, but rather a lack of evidence for a positive decision. Hence, confidence in negative decisions is based on a greater distance in the amount of evidence away from the criterion for a positive response, which constitutes a lack of evidence, rather than on evidence for the type of decision made.

Since confidence in positive decisions is based on the amount of evidence, and confidence in negative decisions is based on lack of evidence, the confidence-accuracy relationship for negative decisions is not likely to be a strong as that for positive decisions. This concept is explained in detail in a later section, and forms the first key part of the explanation proposed later in this chapter.

Considered alone, this disconnection between confidence and evidence for decisions would seem to suggest that the confidence-accuracy relationship for negative decisions will always be poor. However, dual-process theories suggest an additional type of evidence for recognition decisions which may influence how confidence ratings for negative decisions are made. In dual-process models of recognition (e.g., Joordens & Hockley, 2000; Mandler, 1980; Reder et al., 2000; Andrew P. Yonelinas, 1997), the same familiarity process still occurs, but responses and confidence ratings can also be based on a second type of evidence; recollection of the target item which allows match or mismatch of this memory with the test item (e.g., Joordens & Hockley, 2000; Reder et al., 2000). This second type of evidence, recollection, forms the second part of the explanation I proposed for the positive-negative difference. Recollection provides a means for some negative decisions to be made based on evidence for the negative decision, when the recollected item has features that mismatch the test item. Therefore, there is scope for confidence in some negative decisions, where adequate recollected evidence is used, to be based on the degree of mismatch between the test item and the stimulus in memory. Therefore, the second aspect of the explanation I proposed for the positive-negative difference is that it occurs specifically when recollected evidence is not used (adequately) as a basis for the confidence judgment.

Evidence suggests that dual-process models may be more useful for explaining certain types of recognition decisions, particularly complex recognition decisions such as plurality discrimination and associative recognition. Plurality discrimination requires individuals to recognise the subtle difference between words in the form they were studied in, and the same words with changed plurality. Similarly in associative recognition the stimulus to be recognised is not a single item, but an association between two items or contexts, for example, a pair of words, or a face and a name (Hockley & Consoli, 1999; Rotello & Heit, 2000; Weber, Woodard, Williamson, & Hogan, 2014). For these more difficult recognition tasks, use of familiarity alone, as described by single-process theories, seems to be inadequate. Several studies provide evidence that more complex tasks, particularly those where the target item to be recognised has to be distinguished from other very similar items, require the second process described by dual-process theories, of recalling the item that has previously been encountered (Cameron & Hockley, 2000; Giovanello, Verfaellie, & Keane, 2003; Malmberg, 2008; K. A. Norman, 2002; O'Connor, Guhl, Cox, & Dobbins, 2011; Rotello & Heit, 2000). This demonstrates that recollection plays a part in certain recognition decisions. The role it plays in the confidence-accuracy relationship, however, is not yet clear and this relationship is what this thesis aimed to test.

Effect of Recollection on the Confidence-Accuracy Relationship

I argue that the positive-negative difference, specifically the poor resolution of negative decisions, occurs due to neglect of recollection and reliance on familiarity alone. Familiarity alone may not be an accurate indicator of accuracy. This is because confidence in negative decisions increases as familiarity decreases. Thus confidence in negative decisions is based on lack of familiarity which constitutes a lack of evidence. This confounds two factors, one related to accuracy and one not. These two factors both arise from the fact that stimuli that share few features with those in memory will be less familiar. First, unstudied items will, on average, share fewer features with memory than studied items. Therefore, unstudied items will tend to produce higher confidence negative decisions. Thus, considering this factor alone, familiarity appears to provide a genuine basis for confidence as an index of accuracy.[3] Second, familiarity decreases as memory quality becomes poorer. Decreasing memory quality reduces familiarity of all stimuli by reducing the amount of information in memory that matches the test stimulus. Thus, decreased memory quality will result in increased confidence that a negative decision is correct, regardless of whether the stimulus was studied or not. Hence when familiarity is used alone as evidence for a negative recognition decision, memory quality may influence confidence in the opposite direction to accuracy.

On the other hand, for decisions in which recollected evidence is used, the individual has details of the original item studied in mind. This means that confidence can be based on the degree of match (for a positive decision) or mismatch (for a negative decision) between this recollected information and the item shown at test. As memory quality increases, more recollected details are available. Thus as memory quality increases, more matching or more mismatching detail is available on which to base both the recognition decision and confidence. Thus higher memory quality leads to higher match or mismatch and increased confidence. As memory quality decreases, little recollected information is available, resulting in less match or mismatch between the test item and the item in memory, and thus lower confidence. Importantly, match or mismatch does not only provide a more accurate basis for discriminating studied from unstudied items (especially in associative tasks), it causes the basis of confidence to vary in the same way as accuracy. That is, when memory quality is high, unstudied items will produce more mismatch than studied items and, therefore, on average be rejected more confidently. However, as memory quality drops, so too will

[3] It should be noted that the greater the associative nature of a memory task, the less familiarity will discriminate seen from unseen stimuli. In the extreme case, familiarity will not discriminate at all. Consequently, this veridical basis for confidence will decrease in impact as the task becomes more associative in nature.

evidence of mismatch, causing confidence too to drop. Thus, unlike familiarity, factors that cause poor memory (and, therefore lower accuracy) will always cause confidence, on average, to decrease.

While use of familiarity is effective for positive decisions, since very high familiarity is unlikely to occur when an item has never been seen before, very low familiarity may have one of two causes: First, that the test item, and other items like it, have not been seen before, and second, that memory of the target is too poor to produce a match with the test item. In a task such as a lineup, where foils are similar to targets and could be expected to produce some sense of familiarity, very low familiarity is less likely to indicate lack of prior experience with any of the items, and more likely to indicate poor encoding at the time of seeing the original stimulus.

Since the traditional eyewitness lineup and face recognition tasks are forced report (that is, individuals must choose between a positive and a negative response, without the option to say unsure or don’t know) it is possible that negative decisions are sometimes used in the place of a ‘don’t know’ response in these tasks. However, it is also possible that individuals simply neglect recollected evidence when making decisions about faces, and hence confidence for negative decisions suffers the consequences of reliance on familiarity described above. Recognising faces is an ingrained process in humans (Anaki, Nica, & Moscovitch, 2011; Baron, 1981; Bruce & Young, 1986; Jung, Ruthruff, & Gaspelin, 2013) and therefore may be undertaken in the most automatic manner possible. Familiarity is a more automatic and rapid process, and therefore more likely to be used by default for face recognition tasks, as it is suggested to be for item recognition tasks (Malmberg, 2008; Malmberg & Xu, 2007). There is some evidence that recognition memory for faces is not similar to associative recognition for other stimuli, such as word pairs, and tends to instead use a familiarity-based approach (e.g., Rhodes, Castel, & Jacoby, 2008). Therefore it is feasible that individuals engaging in face recognition and eyewitness recognition tasks may neglect recollected evidence. This could explain why the positive-negative difference has been primarily noted in these tasks.

Neglect of recollection as evidence for a decision is a possible explanation for the positive-negative difference. If individuals approach recognition of faces in a manner similar to what is expected for item recognition, they are likely to neglect recollected evidence in these tasks. This should result in patterns of results similar to what is seen in item recognition. However, eyewitness recognition also shares features with associative recognition tasks as the lineup members are deliberately chosen to be similar foils. In addition, some face recognition tasks involve associative information such as names or occupations used to label the faces. Therefore, although it is likely that individuals approach these tasks as they would item recognition, with very little consideration of recollected evidence, it is also likely that use of recollected evidence could improve accuracy, as well as increasing the strength of the confidence-accuracy relationship. Assuming that the relationship between memory quality and confidence described above is common to all recognition tasks, the confidence-accuracy relationship for negative decisions can also be expected to be poor in item recognition tasks where use of familiarity is adequate for overall accuracy and recollection more difficult to use, due to the absence of cues. In addition, restricting the amount of recollection used in tasks like plurality discrimination where it is expected to already be used should negatively impact the confidence-accuracy relationship of negative decisions. Similarly, promotion of recollection in recognition tasks should reduce the positive-negative difference.

The idea that the amount of recollected evidence used in recognition decisions can influence the confidence-accuracy relationship has not yet been thoroughly tested. Some recent evidence from the Flinders lab has supported the idea that use of recollection in a decision has an impact on the confidence-accuracy relationship. Weber et al. (2014) tested participants' recognition of a set of paired general knowledge questions and answers, and compared cases when an answer for the general knowledge question was available in memory to cases where no answer was remembered. The confidence-accuracy relationship for negative recognition decisions where individuals could think of an answer was as strong as that for positive decisions. In contrast decisions where individuals could not recall the original stimulus produced the typical lack of relationship between confidence and accuracy. Hence the confidence-accuracy relationship for negative decisions was improved when recollected information was available during the recognition decision. In addition, the response latency-accuracy relationship was significantly stronger when participants were able to recall an answer, than when they were not. Hence, negative decisions based on knowledge that the presented stimulus did not match participants' memory of the target formed a subset of negative decisions in which confidence and accuracy, and response latency and accuracy, were more strongly related. The overall lack of confidence-accuracy relationship for negative decisions was caused by the second subset of negative decisions, in which participants could not recall the target item and instead made a decision based on the perceived familiarity or appropriateness of the answer presented.

Weber et.al.'s findings appear to disagree with the findings of Sauerland and Sporer (2009) discussed earlier, who found that negative decisions made for different reasons did not demonstrate any differences in the confidence-accuracy relationship. In contrast the results of Weber et.al.'s study suggest that negative decisions made for different reasons may produce different relationships between accuracy and its possible predictors. The discrepancy in results could be because Sauerland and Sporer's distinction between reasons for making negative decisions was not clear enough to participants, or because other factors caused the difference between instances where participants could recall and answer and could not recall an answer in Weber et. al.'s study, which was quasi-experimental. One possible influence on Weber et. al.'s results could be memory quality as this could allow participants to recollect an answer and also assist them to make a better informed recognition decision. On the other hand evidence suggests that individuals are not adept at determining the reasons behind their recognition decisions or at understanding the distinctions experimenters make between decisions based on recollection or familiarity (e.g., Dunn, 2004; Mickes, Seale-Carlisle, & Wixted, 2013), and hence Sauerland and Sporer's instructions may not have clearly distinguished decisions involving recollection from those involving familiarity or guessing. This thesis aimed to provide experimental evidence to determine whether the evidence bases used for recognition decisions do influence the confidence-accuracy relationship. Specifically I suggested that neglect of recollected evidence might produce the positive-negative difference since familiarity-based confidence ratings for negative decisions do not have a strong evidence base, as described earlier. In addition, since Sauerland and Sporer's published study had investigated possible reasons for the positive-negative difference specifically in the eyewitness context, this thesis aimed to extend investigations such as Weber et. al.'s which aim to investigate whether the positive-negative difference is general to a variety of tasks in which use of recollection may vary.

Both Sauerland and Sporer's and Weber et. al.'s studies inquired about recollection in categorical terms. In Sauerland and Sporer's study participants were asked to categorically respond with which reason was behind their negative decision, and in Weber et. al.'s study decisions were classified as either having recollected evidence available or not based on what participants could later recall. Recently dual-process theories take into account both the dual-process and the signal-detection accounts by assuming that recollection does occur, as well as familiarity, but is on a continuous scale. The most recent of these is the continuous dual process signal detection model (CDP) proposed by Wixted and Mickes (2012). In CDP, all stimuli trigger a certain level of familiarity and recollection, and these are combined to produce an overall quantity of evidence for a positive decision. If this exceeds the individual's criterion for a positive response, then they consider the item to be previously seen. However, individuals are aware of the origin of this evidence and if asked can also rate to what extent they recollect details of seeing the item and to what extent they feel familiar with it. Recollection is a continuous value, and therefore can influence the combined value in a negative direction if the recollected evidence weighs against the item having been seen before. Negative decisions occur when the combined 'strength' of familiarity and recollection evidence is inadequate to support the item having been seen before.

If recollection is continuous, as Wixted and Mickes suggest, Sauerland and Sporer's categorical questions about the reason for participants' negative decisions may not have clearly identified where recollection was used, and this could pose another reason for their results. In the case of Weber et. al.'s study, whether or not participants could recall an answer after giving a recognition response may not always indicate whether this recollected information was used or available during the recognition decision. While the experiments in this thesis do not attempt to manipulate recollection on a continuous scale, it is assumed that the different experimental conditions make either more or less recollection available for recognition decisions rather than making recollection available or unavailable.

Before describing the types of experiments conducted however, it is important to note the reasons for investigating the confidence-accuracy relationship for negative decisions specifically, as many of the established problems with tasks such as eyewitness identification are normally framed in terms of the danger of incorrect positive decisions. If the positive-negative difference occurs in other recognition tasks depending on the level of recollection used as I suggested, there are a series of other contexts which an investigation of the relationship between recollection and confidence-accuracy may shed light on.

Importance of Understanding the Positive-Negative Difference

There are a number of contexts in which incorrect recognition decisions can have severe consequences. Therefore it is desirable to be able to predict the accuracy of an individual's decision. In addition, understanding the cause of the positive-negative difference has theoretical benefits for the understanding of recognition memory in general. For example, the debate over single-process versus dual-process theories of recognition memory is ongoing. If support was found for the idea that recollected evidence produces changes in the confidence-accuracy relationship, producing or reducing the positive-negative difference, this provides further evidence for dual-process models of recognition as only these models would so far offer an explanation for the positive-negative difference. If the positive-negative difference also varies in size based on the amount of recollected evidence available for decisions, and is not merely present or absent, this would also suggest that recollection is continuous as suggested by Wixted (e.g., Wixted & Mickes, 2010; Wixted & Stretch, 2004), not an all-or-nothing threshold process as suggested by some types of dual-process theories. On the other hand if the positive-negative difference only occurs in situations where recollection is impaired, this would provide evidence for a threshold recollection process where recollection is only used for specific recognition decisions.

Another theoretical point to which an investigation of this idea may contribute is how individuals determine whether to include recollected evidence in their recognition decisions. If as Malmberg (2008) suggests, individuals choose whether to include recollected evidence based on how efficient its use is for specific types of decisions, it should be possible to encourage individuals to use recollection more by changing the degree to which they view it as efficient, or the degree to which they feel it is necessary to make rapid decisions. On the other hand if recollection is only possible for certain tasks, it should not be possible to alter the degree of recollection used in decisions by changing individuals' desire to use recollected detail. Hence increasing understanding the connection between recollection and the confidence-accuracy relationship can provide evidence towards the resolution of a number of contested theoretical issues.

In addition there are multiple practical issues which knowledge of this issue may suggest solutions for. Owing to the fact that the positive-negative difference was first noted in eyewitness identification studies, past research has often focussed on the serious outcomes of incorrect recognition decisions in the context of eyewitness identification. The vast majority of research in the eyewitness field focuses on the outcomes of positive decisions (e.g., Dobolyi & Dodson, 2013; " Innocence Project. ," 2013; Wells, 2014). Although it is inaccurate positive decisions by eyewitnesses (false alarms) that can result in convictions of innocent individuals, inaccurate negative decisions (misses) may also have serious consequences. Failure of an eyewitness to identify a criminal may result in police considering it less likely that the suspect committed the crime, dedicating less attention to investigating that individual and instead searching for an alternative suspect. If the original suspect is the culprit, this may result in delayed investigations and even opportunity for the culprit to commit further offenses. Since police are more likely to believe a witness who rejects the lineup confidently (Boyce, Lindsay, & Brimacombe, 2008), a witness who behaves in this way may be awarded more credibility, despite the fact that confidence currently has no bearing on the accuracy of negative decisions. If considered from the reverse direction, being able to determine the accuracy of a negative decision is also important so that a suspect can be exonerated on the basis of a witness's negative decision. Police may otherwise take a lineup rejection as a lack of evidence rather than evidence that the suspect is innocent, which a rejection may provide if it were possible to estimate its accuracy. Mathematical modelling suggests that lineup rejections should provide evidence that the suspect is innocent, but information on real-world cases suggests that rejections are often instead viewed as a lack of evidence (Wells & Lindsay, 1980).

In other recognition contexts an incorrect negative decision may be the more problematic error, for example, if a doctor incorrectly determines that there are no abnormal cells present on a microscope slide which does show abnormal cells. Therefore, despite the focus on false identifications in the eyewitness literature, the examination of negative decisions and the explanation for the positive-negative difference is critical to understanding the recognition memory process and to real world applications of recognition memory theory. It is hoped that testing the explanation proposed in this thesis will identify methods of improving the relationship between confidence and accuracy for negative decisions, so that these markers may be used to assist in determining the accuracy of all decision types. In addition, understanding the cognitive processes that lead to the positive-negative difference will enhance our understanding of recognition memory processes in general and contribute to a consensus on which of the available models of recognition memory best explains these processes.

Rationale for the Experiments

One method that has been used to investigate the role of recollection in recognition decisions is to attempt to prevent it occurring in some cases and observe the effect on various outcome variables. However the effect on the confidence-accuracy relationship is yet to be investigated in this way. A common point made by dual-process theories is that familiarity evidence accumulates rapidly while recollected evidence is slower to become available (Basile & Hampton, 2013; Malmberg, 2008; A. P. Yonelinas & Jacoby, 1994). Partly as a result of this some theorists (e.g., Malmberg, 2008) suggest that individuals may regulate whether they use recollected evidence depending on what they view as efficient for the task.

Individuals select which strategy they use (faster familiarity only or slower decisions including more recollected evidence) according to which is most 'efficient' in the situation for achieving their goals, and for the type of task. Familiarity is usually sufficient for item recognition, but in situations where targets must be discriminated from similar foils, recollection becomes vital to accuracy. Many studies which provide evidence that dual-process models best describe the recognition process have used the approach of speeding decisions (e.g., Jacoby et al., 1998; Jones & Jacoby, 2001; McElree, Dolan, & Jacoby, 1999; Odegard, Koen, & Gama, 2008; Sauvage, Beer, & Eichenbaum, 2010; A. P. Yonelinas & Jacoby, 1994), and demonstrated that recollected evidence does take longer to accumulate. Hence an individual who is concerned more with accuracy than speed will wait for recollected information to become available, but someone who is concerned with the speed of a decision or views the familiarity evidence as strong enough on its own may neglect recollected evidence. This means that individuals may make an inappropriate choice to rely on familiarity in order to respond rapidly, particularly when familiarity is exceptionally low, or when doing a task which they perceive as not requiring recollected information. Hence it is possible that in eyewitness identification and face recognition, individuals expect familiarity to provide adequate evidence, and do not wait for recollected evidence to become available. This could produce neglect of recollection in these tasks which in turn may result in a positive-negative difference for the reasons described earlier.

The aim of the experiments described in this thesis was to examine whether there were changes in the confidence-accuracy relationship for negative decisions in laboratory recognition tasks depending on the extent to which recollection was able to be used. If neglect of recollection underlies the positive-negative difference in face recognition and eyewitness identification, differences in recollection in other recognition tasks should have effects on the confidence-accuracy relationship. In the first experiments, an item recognition task, in which use of recollection could be expected to be low, was compared with plurality discrimination, in which the use of recollection is expected to be higher by comparison (Malmberg, 2008). If the explanation I proposed for the positive-negative difference is accurate, the item recognition task could be expected to demonstrate a positive-negative difference similar to that normally observed in face recognition and eyewitness recognition. This is because item recognition is not considered to involve any substantial amount of recollection, and my explanation hinges on recollection being neglected in tasks which demonstrate the positive-negative difference. In the subsequent experiments recollection was manipulated, first with manipulations that were expected to impair recollection, and finally with a manipulation to increase recollection. This was done in tasks which normally do involve recollection, such as plurality discrimination. If the amount of recollection used was the key to the poor confidence-accuracy relationship for negative decisions in item recognition and face recognition tasks, I expected to see a similarly poor confidence-accuracy relationship in associative tasks when recollection was impaired. In contrast, when recollection was encouraged I expected to see a strengthened confidence-accuracy relationship for negative decisions.

In summary, recognition decisions can have important consequences in some situations, and as such identifying variables which may help diagnose their accuracy has a variety of practical applications. One major variable which has been investigated as a predictor of recognition accuracy is confidence in the response given. Although confidence has shown some evidence of predicting accuracy in recognition research more generally, in face recognition and eyewitness identification it has frequently been found to predict accuracy for positive, but not negative decisions. The aim of the experiments conducted in this thesis was to test the explanation that this difference between positive and negative decisions is caused by neglect of recollected evidence.

References

Anaki, D., Nica, E. I., & Moscovitch, M. (2011). Automatic aspects in face perception evidence from mandatory processing of distractor facial components. *Experimental Psychology*, *58*(1), 4-18. doi: 10.1027/1618-3169/a000061

Balakrishnan, J.D. , & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of experimental psychology. Human perception and performance*, *22*(3), 615-633.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics, 55*(4), 412-428. doi: 10.3758/bf03205299

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of experimental psychology. Human perception and performance*, *24*(3), 929-945.

Baron, Robert J. (1981). Mechanisms of human facial recognition. *International Journal of Man-Machine Studies, 15*(2), 137-178. doi: http://dx.doi.org/10.1016/S0020-7373(81)80001-6

Basile, B. M., & Hampton, R. R. (2013). Recognition errors suggest fast familiarity and slow recollection in rhesus monkeys. *Learning & Memory, 20*(8), 431-437. doi: 10.1101/lm.029223.112

Boutis, K., Pecaric, M., Seeto, B., & Pusic, M. (2010). Using signal detection theory to model changes in serial learning of radiological image interpretation. *Advances in Health Sciences Education, 15*(5), 647-658. doi: 10.1007/s10459-010-9225-8

Boyce, M., Lindsay, D. S., & Brimacombe, C. (2008). Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators. *Law and Human Behavior, 32*(5), 439-453. doi: 10.1007/s10979-007-9125-5

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22*(6), 827-840. doi: 10.1002/acp.1486

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science, 23*(10), 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G.L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of experimental psychology. Applied, 12*(1), 11-30.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Cameron, T. E., & Hockley, W. E. (2000). The revelation effect for item and associative recognition: Familiarity versus recollection. *Memory & Cognition, 28*(2), 176-183. doi: 10.3758/bf03213797

Cary, M. , & Reder, L.M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49*(2), 231-248.

Clark, S. E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of experimental psychology. Learning, memory, and cognition, 23*(1), 232-238.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*(1), 37-60. doi: 10.3758/bf03210740

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of experimental psychology. Learning, memory, and cognition, 27*(2), 414-435.

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology-Applied, 19*(4), 345-357. doi: 10.1037/a0034596

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*(2), 524-542. doi: 10.1037/0033-295x.111.2.524

Festinger, L. (1943). Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology, 32*(4), 291-306.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review, 91*(1), 1-67.

Giovanello, K. S., Verfaellie, M. , & Keane, M. M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cognitive, Affective, & Behavioral Neuroscience, 3*(3), 186-194. doi: 10.3758/cabn.3.3.186

Hardweier, D., Hofer, F., & Schwaninger, A. (2005). The X-Ray object recognition test (X-Ray ORT) - A reliable and valid instrument for measuring visual abilities needed in X-Ray screening. In L. D. Sanson (Ed.), *39th Annual 2005 International Carnahan Conference on Security Technology, Proceedings* (pp. 189-192). New York: Ieee.

Heathcote, A., Bora, B., & Freeman, E. (2010). Recollection and confidence in two-alternative forced choice episodic recognition. *Journal of Memory and Language, 62*(2), 183-203. doi: 10.1016/j.jml.2009.11.003

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review, 95*(4), 528-551.

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27*(4), 657-664. doi: 10.3758/bf03211559

Innocence Project. . (2013). 2014, from http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php

Jacoby, L. L. (1991). A process dissociation framework - separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. doi: 10.1016/0749-596x(91)90025-f

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*(4), 705-709. doi: 10.3758/bf03208849

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*(1), 82-102. doi: 10.1006/jmla.2000.2761

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(6), 1534-1555. doi: 10.1037//0278-7393.26.6.1534

Jung, K., Ruthruff, E., & Gaspelin, N. (2013). Automatic identification of familiar faces. *Attention Perception & Psychophysics, 75*(7), 1438-1450. doi: 10.3758/s13414-013-0468-3

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. (1971). Recognition time for informations stored in long-term memory. *Perception & Psychophysics, 10*(1), 8-&. doi: 10.3758/bf03205757

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 15*(6), 659-671. doi: 10.1002/acp.739

Lichtenstein, Sarah, & Fischhoff, Baruch. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159-183. doi: http://dx.doi.org/10.1016/0030-5073(77)90001-0

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9*(3), 215-218.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*(3), 545-556. doi: 10.3758/bf03193293

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review, 87*(3), 252-271.

McCarley, J. S. (2009). Effects of speed-accuracy instructions on oculomotor scanning and target recognition in a simulated baggage X-ray screening task. *Ergonomics, 52*(3), 325-333. doi: 10.1080/00140130802376059

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of*

Experimental Psychology-Learning Memory and Cognition, 25(3), 563-582. doi: 10.1037//0278-7393.25.3.563

Mickes, Laura, Seale-Carlisle, Travis M., & Wixted, John T. (2013). Rethinking familiarity: Remember/know judgments in free recall. *Journal of Memory and Language, 68*(4), 333-349. doi: http://dx.doi.org/10.1016/j.jml.2013.01.001

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological review, 89*(6), 609-626.

Murdock, B. B., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology, 94*(3), 284-290.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6*(2), 192-208. doi: 10.1016/0022-2496(69)90002-9

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology-Learning Memory and Cognition, 28*(6), 1083-1094. doi: 10.1037//0278-7393.28.6.1083

O'Connor, A. R., Guhl, E. N., Cox, J. C., & Dobbins, I. G. (2011). Some memories are odder than others: Judgments of episodic oddity violate known decision rules. *Journal of Memory and Language, 64*(4), 299-315. doi: 10.1016/j.jml.2011.02.001

Odegard, T. N., Koen, J. D., & Gama, J. M. (2008). Process demands of rejection mechanisms of recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(5), 1296-1304. doi: 10.1037/a0013034

Ratcliff, R. (1978). Theory of memory retrieval. *Psychological Review, 85*(2), 59-108. doi: 10.1037//0033-295x.85.2.59

Reder, L.M., Angstadt, P., Cary, M., Erikson, M.A. , & Ayers, M.A. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Learning, memory, and cognition, 26*, 294-320. doi: 10.1037/0278-7393.26.2.294

Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and aging, 23*(2), 239-249. doi: 10.1037/0882-7974.23.2.239

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*(6), 907-922. doi: 10.3758/bf03209339

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34*(4), 337-347. doi: 10.1007/s10979-009-9192-x

Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior, 36*(5), 394-403. doi: 10.1037/h0093926

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of experimental psychology. Applied, 15*(1), 46-62.

Sauvage, M. M., Beer, Z., & Eichenbaum, H. (2010). Recognition memory: Adding a response deadline eliminates recollection but spares familiarity. *Learning & Memory, 17*(2), 104-108. doi: 10.1101/lm.1647710

Schwark, J., Sandry, J., MacDonald, J., & Dolgov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention Perception & Psychophysics, 74*(8), 1583-1589. doi: 10.3758/s13414-012-0354-4

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145-166. doi: 10.3758/bf03209391

Slotnick, S. D. (2010). Remember source memory ROCs indicate recollection is a continuous process. *Memory, 18*(1), 27-39. doi: 10.1080/09658210903390061

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22*(2), 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of applied psychology, 78*(1), 22-33.

Sporer, S. L., Penrod, S. D., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological bulletin, 118*(3), 315-327.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of experimental psychology. Learning, memory, and cognition, 26*(3), 582-600.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of applied psychology, 88*(3), 490-499.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of experimental psychology. Applied, 10*(3), 156-172.

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*(1), 17-31. doi: 10.1002/acp.1166

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: the unruly 10-12-second rule. *Journal of experimental psychology. Applied, 10*(3), 139-147.

Weber, N., & Varga, M. (2012). Can a modified lineup procedure improve the usefulness of confidence? *Journal of Applied Research in Memory and Cognition, 1*(3), 152-157. doi: http://dx.doi.org/10.1016/j.jarmac.2012.06.007

Weber, N., Woodard, L., Williamson, P., & Hogan, A. (2014). *How do we recognise what we haven't seen? Different bases for positive and negative decisions.* Manuscript submitted for publication.

Wells, G. L. (1993). What do we know about eyewitness identification? *The American psychologist, 48*(5), 553-571.

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science, 23*(1), 11-16. doi: 10.1177/0963721413504781

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness non-identifications. *Psychological Bulletin, 88*(3), 776-784. doi: 10.1037/0033-2909.88.3.776

Wixted, J. T., & Mickes, Laura. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616-641.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of experimental psychology. Learning, memory, and cognition, 20*(6), 1341-1354.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(6), 1415-1434. doi: 10.1037/0278-7393.25.6.1415

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory - effects of interference and of response speed. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 48*(4), 516-535. doi: 10.1037/1196-1961.48.4.516

Yonelinas, Andrew P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*(6), 747-763. doi: 10.3758/bf03211318

CHAPTER 2 — ANALYTICAL APPROACHES

The purpose of this chapter is to detail the approaches used to analyse the results of the experiments in subsequent chapters, and explain their value for analysing the type of data yielded by the experiments conducted. Traditional approaches to data analysis encounter a number of problems when dealing with data from experiments containing multiple items and multiple memory tests. Therefore data analysis in this thesis used a mixed effects modelling approach as explained below.

Problems with Traditional Analyses for Experiments with Multiple Items

In the experiments reported in this thesis, each participant studied multiple items, and provided multiple responses, confidence ratings, response latencies, and decision types. For this type of data, traditional approaches to data analysis encounter three major issues. The first major problem is that analyses must be based on aggregate statistics such as proportion correct. Aggregate statistics such as the mean proportions correct, mean hit rate, mean confidence, and so on can be misleading because they may hide effects that are present or make it appear as though effects are there that are not (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). Taking confidence ratings as an example, participants can get the same mean confidence by using confidence ratings that have very different ranges, for example, a participant who only uses the 70% and 90% ratings versus someone who only uses 60% and 100%. These two participants would have the same mean confidence (80%), however this mean confidence represents the first participant much better than the second. In addition, one individual can have a mean confidence of 80% by only using the 80% rating, while another may use the full confidence scale, yet still have a mean confidence of 80%. If the second individual gives lower confidence ratings to incorrect decisions and higher ratings to correct decisions then they should display a superior confidence-accuracy relationship. However, their mean confidence would be the same as the individual who only used the 80% option, and therefore any index of the confidence-accuracy relationship calculated from the mean confidence would fail to identify the large difference between these individuals.

The second major problem with traditional data analysis methods is that they cannot take into account variation in the effects of manipulations between stimulus items and between individual participants at the same time. All of the experiments in this thesis used samples of word stimuli (for example, nouns between 4 and 10 letters in length, chosen at random from a list of words that fall within certain ranges on imagability, concreteness and written frequency). Such samples do not exhaust the list of all possible words that could be used as stimuli. More importantly the different stimuli in the selected sample may vary in their memorability. In addition the characteristics of these samples of words may differ from samples of words used in other experiments. Significant variation may exist in the effects in question (for example variation in the ability of confidence to predict accuracy) between stimulus items. More memorable words may produce stronger confidence-accuracy relationships, for example, due to being more strongly encoded. This variation in effects between stimulus items has been found to inflate Type 1 error rates in traditional analyses. Such analyses calculate meta-memory accuracy such as the confidence-accuracy relationship of interest here for each participant, but do not take into account any differences in the effects being studied between stimulus items (Murayama, Sakaki, Yan, & Smith, 2014).

The third problem with aggregate measures of recognition accuracy is that often these measures for individual participants can only be calculated if the participant gives responses in each category. For example, d prime can only be calculated (without adjustment) for a participant whose responses include both hits and false alarms, and who also does not have a 100% hit rate, 100% false alarm rate, 0% hit rate or 0% false alarm rate. This often results in either adjustments being applied to the data of participants who obtain 100% hits or 0% false alarms, such that it does not fully represent their responses, or in those participants who do not give responses in a certain category being dropped from analyses (Murayama et al., 2014). This means that the portion of the data provided by these participants cannot be used when conducting analyses using aggregate measures.

Mixed Effects Modelling

An alternative to traditional data analysis methods which does not use aggregate statistics for each participant, and hence better represents all of the data is mixed effects modelling. This approach provides a solution to all of the problems with use of aggregate statistics calculated by participant discussed above. Each data point from each participant contributes to the model, rather than an aggregate measure being created for each individual or for each question, thereby more accurately representing the full scope of responding, and avoiding problems with aggregate measures hiding differences between participants. In addition all data can be used, rather than participants whose responses do not cover all the necessary categories for calculation of aggregate statistics being excluded. Importantly the effects of differences between stimuli can be explicitly accounted for in the model. This avoids inflation of Type 1 error (Judd, Westfall, & Kenny, 2012; Murayama et al., 2014). The sections below will explain in more detail how this is achieved.

Mixed effects modelling is an extension of regression, in that a model is created to estimate the change in an outcome variable with changes in a predictor variable. However it differs from a normal regression in that rather than having just one error term which is assumed to account for all variation not caused by the predictor, part of the error variance is explicitly modelled as the changes caused by variation between individuals or items. The name mixed effects reflects the fact that these models combine fixed effects and random effects. Fixed effects are factors for which every variation in the population is represented in the experiment, such as gender and experimental manipulations. Random effects are factors for which only a sample of the entire population is used, such as specific participants and words chosen to be used as stimuli (Bates, 2008). Variation between participants and stimuli that does not reflect the effect of the variables being investigated can be accounted for by entering these factors (participant and stimuli) as random effects, allowing either the intercept or intercept and slope to vary randomly for these effects. When a variable (e.g., stimulus item) is entered as a random effect, the model will estimate the variability (indexed as standard deviation) of the outcome variable associated with changes in the random effect (in stimulus item). On a conceptual level, this is achieved by the model calculating each parameter separately for every level of the random effect (for every stimulus item). In other words, the model calculates a different intercept coefficient (which indicates the base level of performance when the impact of other predictors is ignored) or slope coefficient (which indicates the amount of change in the outcome variable with a one unit change in a predictor) for each level of each random effect. Hence by entering stimulus as a random slope[4] and intercept in all models, I took into account the fact that the effects of interest could vary by

[4] Random slopes are only appropriate when each level of the random effect occurs in every level of the fixed effect. For example, in the experiments in this thesis the same word list was used in each condition, so each word appeared in both conditions; thus stimulus word is included as a random slope. If different word lists were used for the different conditions, however, the effect of condition could not vary based on stimulus word, as each condition would only apply to one word list. Thus including stimulus word as a random effect would not be appropriate. The appropriateness of including participant as a random slope and intercept depends on study design. In between-subjects designs a different set of participants complete the different conditions. Therefore differences between participants contain the effect of interest; the difference between conditions. Thus inclusion of participant as a random effect is not appropriate. For within-subjects designs, the same participants complete both conditions, and hence the effect of condition may differ for different participants. Hence inclusion of participant as a random effect is appropriate.

stimulus item. Participant number was also included as a random slope and intercept for within-subjects analyses. In within-subjects designs, the effect of a manipulation may vary by participant, and thus including participant as a random effect is appropriate. In a between-subjects design, different groups of participants complete the different levels of the manipulation. This means that differences in the effect of the manipulation between participants do not reflect differences in participants in how they respond to the manipulation, but differences between conditions. Thus it is only appropriate to include participant as a random effect for within-subjects comparisons.

Mixed effects modelling also allows simultaneous consideration of multiple predictor variables in the same analysis. In my analyses I was interested in how the relationship between confidence and accuracy changed depending on experimental conditions and also on the type of response (positive or negative) given. I was also interested in whether a positive-negative difference was still present, reduced or eliminated following each of the manipulations, so I wanted to factor in all of the predictor variables and their different interactions in the same analysis. Using mixed effects modelling I was able to do this to determine whether response, experimental condition and confidence interacted in different combinations in their relationship with accuracy. Although more traditional analyses such as ANOVAs do allow consideration of multiple predictors, the fact that effects of stimuli and participants are not explicitly modelled in the error variance means that as more predictors are added the error variance must be assumed to increase and small effects may be hidden. In mixed effects models because some of the error variance is explicitly accounted for as due to variations between participants and stimuli the increase from adding more predictors is reduced. In addition calculating aggregates to represent individuals' confidence-accuracy relationships, as is necessary in order to conduct other types of analyses, may exaggerate the effect of outliers in stimuli or within the data of individuals.

For each analysis in this thesis, mixed effects models were created using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2014) in `R`, an open-source language environment for statistical computing (R Core Team, 2014). The outcome in all models was either false alarms (binomial, where false alarm is 1 and correct rejection is zero), hits (binomial, where hit is 1 and miss is zero) or accuracy (binomial, where correct is 1 and incorrect is zero). Hence the outcome variables were all dichotomous and a logit link function was most appropriate, making the mixed effects modelling an extension of logistic regression rather than normal regression. The distinguishing feature of logistic regression is that rather than predicting the outcome variable from a predictor variable as in linear regression, the probability of the outcome given a certain value of the predictor is estimated (Jaeger, 2008).

In each analysis I first fitted a baseline model which included the intercept and random effects (stimulus (word) and for within-subjects experiments, participant number) to the data. Predictors of interest were then added step by step in order of those of least interest that I wanted to control for, progressing up to the highest-order interaction of interest. The model generated a regression coefficient for each added predictor (*b),* which in logistic regression represents the amount of change in the log odds of the outcome with one unit change in the predictor, when all other variables are zero. For the dichotomous binomial predictors such as decision type (positive or negative), the *b* value represents the amount of change in the slope of the line of best fit between the 0 and 1 values of the predictor. For example, the *b* value for the effect of response (positive or negative) would represent the amount of change in the line of best fit between positive decisions (0) and negative decisions (1) for when all other predictors are zero.

For mixed effects models, there are a number of ways in which significance tests can be conducted. For each factor added to the model, the extent to which the fit of the model is improved can be tested. However, this does not directly assess the effect of the factor of interest in the model. To more directly test whether the coefficients representing the interactions of interest differed significantly from zero, I used 95% confidence intervals to determine whether the *b* values were significantly different from zero. Effects for which the 95% confidence interval did not include zero (alpha level < = .05) were interpreted as significant. Confidence intervals were calculated using bootstrapping with the `boot` package in `R` (A. Canty & Ripley, 2014). As the distributional properties of the estimates given in logistic mixed effects models are unknown, significance tests which assume normal distributions cannot be used. For this reason, bootstrapped confidence intervals were used instead. Bootstrapping takes random samples with replacement from the original data to create a set of N data sets. For each of the resulting data sets, the analysis is then run, calculating the coefficients. The distribution of the coefficients over all the data sets can then be used to estimate the standard error of the coefficient (*b* estimate) for the sample in the experiment and directly estimate a confidence interval for the *b* value. For all analyses in this thesis the confidence intervals presented are based on 1000 bootstrap samples. This number of samples was used for practical reasons as 1000 samples is adequate to give a stable estimate of the confidence interval and can be calculated in a day. Small increases in this stability which could be obtained from using more samples are impractical to obtain due to the amount of time it would take to calculate the intervals, as the time to run the analyses increases dramatically with only a small gain in the precision of the estimates (Davison & Hinkley, 1997).

To further explain the mixed effects modelling approach consider an example from Chapter 3. Table 1 shows the model coefficients from experiment 1a. This experiment was an item recognition task with words. The analysis is testing for the presence of a positive-negative difference. Since the positive-negative difference is a difference in the confidence-accuracy relationship for positive and negative decisions, it can be reflected as an interaction between decision type (positive, negative) and confidence on the outcome of accuracy. The model which produced the statistics in Table 1 analysed this interaction. For each factor the

Table 1

Fixed Effects Coefficients for Confidence and Accuracy in Study1 by Response Type

Fixed Effect, level	*b*	SE_b	95% CI[b]
	Confidence and accuracy		
Intercept	-0.74	0.16	[-0.98, -0.53]
Response, Negative (R)	1.34	0.20	[0.97, 1.74]
Confidence (C)	0.52	0.05	[0.44, 0.61]
R × C	-0.39	0.06	[-0.51, -0.27]

Note. The models included participant (*SD* = 0.17) and question (*SD* = 0.32) as random slopes and intercepts. [a] Confidence responses were elicited on a 50% to 100% response scale but coded in the analysis as 0 through 5 (i.e., the percentage minus 50, then divided by 10). [b] Confidence intervals were calculated on 1,000 bootstrap samples using the first order normal approximation implemented in the boot package for R (Angelo Canty & Ripley, 2013).

table lists the regression coefficient (*b)* the standard error for that coefficient, and the 95% confidence interval for the coefficient. Recall that the *b* value represents the change in the log odds of the outcome (accuracy) for one unit change in the predictor when all other variables are zero. In all datasets throughout this thesis, decision type was coded with positive decisions as 0 and negative decisions as 1. Therefore the model uses positive decisions as the baseline category and reflects the change between positive and negative decisions. Hence the positive regression coefficient of 1.34 for the effect of response indicates that if confidence were zero[5], negative decisions (coded as 1) would be more likely to be accurate than positive decisions. However, this does not provide much information as it only concerns one confidence category; the lowest. The coefficient of interest for answering the question is that for the interaction. Note that the effect of confidence (when response is zero, that is, a positive decision) is positive, indicating that increasing confidence is associated with increasing accuracy. Therefore the negative *b* value for the interaction indicates that when the response was a negative decision (1) rather than a positive decision (0), the positive effect of increasing confidence on accuracy was smaller. Hence the positive relationship between confidence and accuracy was smaller for negative than positive decisions, indicating a positive-negative difference. The 95% confidence interval for the interaction between confidence and response does not include zero, indicating that within the possible datasets produced by the bootstrapping process, the coefficients were consistently negative. This can be considered significant at $\alpha = .05$

Other Methods of Quantifying the CA Relationship

There are a number of other methods of measuring the confidence-accuracy relationship which have been used in past research, which I will explain here for comparison purposes. Past approaches to analysing the confidence-accuracy relationship in recognition have included calculation of calibration (and associated over and under-confidence) or resolution,

[5] Although participants rated their confidence on a scale of 50-100%, these values were recorded in some of the datasets (including this example) as 0-5, respectively. Therefore, when confidence is zero this reflects responding in the 50% category. This was not the case for all data sets in this thesis; some recorded confidence as rated by participants ranging from 50 to 100. In these cases, statistics which refer to the 0 category for confidence are uninformative.

as indexed either by correlation or the Normalised Resolution Index (ANRI) of confidence and accuracy (e.g., Brewer, Keast, & Rishworth, 2002; Olsson, 2000). Early studies in the eyewitness recognition field which used correlation to quantify the confidence-accuracy relationship found very low correlations (Olsson, 2000). Subsequently authors suggested that correlation was not a good index of the confidence-accuracy relationship and advocated for the use of calibration and different statistics to index resolution or discrimination (Juslin, Olsson, & Winman, 1996; Olsson, 2000).

Calibration and resolution reflect different properties of the confidence-accuracy relationship. Resolution describes the more intuitive aspect; whether confidence can be used to discriminate correct from incorrect decisions. It can be quantified by a number of statistics including point-biserial correlation. However use of point-biserial correlation has been criticised as placing unrealistic expectations on confidence as an indicator of accuracy. This will be explained in more detail along with some alternative measures in the next section. Calibration on the other hand reflects the extent to which individuals' confidence ratings are realistic representations of the likelihood that their decision is correct. That is, calibration indexes the degree to which confidence ratings given as probabilities are representative of the probability that the decision is correct. For example, if 50% of decisions made with 50% confidence are correct, 60% of decisions made with 60% confidence are correct, and so on, perfect confidence-accuracy calibration would be shown. A number of statistics can also be used to quantify calibration, which I will describe briefly below.

Specifics of Statistics used in the Past to Quantify the CA Relationship

There are a number of statistics that have been used to quantify resolution. Resolution indicates the degree to which confidence discriminates correct from incorrect decisions. For example, if all incorrect decisions were given a lower confidence rating than all correct decisions, perfect resolution would occur. The more the confidence ratings for correct and incorrect decisions overlap, the less resolution is shown. This can be indexed by the Normalised Resolution Index (NRI; and better the Adjusted Normlised Discrimination Index ANRI, also refered to as ANDI using discrimination in place of resolution) which ranges from 0 (no discrimination)to 1 (complete discrimination). It is also represented by point-biserial correlation coefficients. However to obtain perfect discrimination, individuals would have to put all confidence estimates for correct decisions in one category, and all confidence estimates for incorrect decisions in another, single category. Therefore use of the correlation coefficient to index relationship strength has been suggested to put unrealistic expectations on the confidence-accuracy relationship as individuals are unlikely to use only two confidence categories to represent their confidence in memory, although they may still rate incorrect decisions less confidently than correct decisions.

Calibration can be represented by the calibration statistic (*C*), with 0 indicating perfect calibration and 1 no calibration. Calibration is often displayed graphically by drawing calibration curves. The perfect calibration line (where 50% of decisions made with 50% confidence are correct, 60% of decisions made with 60% confidence and so on) forms a straight diagonal, and it is possible to see how individuals' responses vary from this line by plotting the mean proportion correct in each confidence category over this line. The calibration statistic, mentioned above, is an index of how far calibration for a sample lies from the perfect diagonal, thus explaining why zero is equal to perfect calibration. Finally calibration can be represented by the over/under confidence statistic, which is intuitive to understand with positive values representing over-confidence and negative values under-confidence.

A problem with both calibration and resolution statistics is that like other traditional analyses, they require aggregate statistics to be calculated for each individual and then for the whole sample for each confidence category, potentially hiding variation within individuals and items as explained earlier. In the analyses presented in this thesis indexes of resolution and calibration are replaced by the *b* values and confidence intervals for the effect of confidence. The coefficients in the models are similar to an estimate of resolution. They indicate the extent to which the probability of accuracy changes as confidence increases (e.g., the change in the log odds of an accurate decision from when confidence is zero (50%) to when it is 1 (60%). However they allow the analysis to use all of the data points provided by each participant to produce the model, rather than calculating aggregate measures.

Estimating Changes in Recollection

A second task for the analyses in this thesis was to investigate the types of recognition errors made to estimate the extent to which recollection was used under different conditions. The type of errors individuals make can reveal the extent to which they are using familiarity evidence versus recollected evidence. False alarms to similar foils are more likely to occur when individuals base their decisions on familiarity, and respond to the familiar features of similar items (e.g., Jacoby, Jones, & Dolan, 1998). The norm for investigating the extent to which this occurs is to calculate false alarm and hit rates (e.g., Higham, Perfect, & Bruno, 2009; Hockley, 2008; Jacoby et al., 1998; Xu & Malmberg, 2007). Parallel increases in the hit rate and false alarm rate are viewed as indicative of low use of recollection, as increasing familiarity is expected to result in increased positive responses. Increases in the hit rate without corresponding increases in the false alarm rate indicate high use of recollected evidence, as they reflect increasing familiarity being offset for similar foils by recollection of contrasting details of the remembered stimulus (Jacoby et al., 1998). An increase in the hit rate accompanied by a decrease in the false alarm rate is known as a mirror effect and occurs when both recollection and familiarity are increased (Reder et al., 2000). Corresponding changes the hit and false alarm rate are sometimes used to calculate *d* prime, an indicator of participants' ability to discriminate between studied and unstudied items, which should increase when recollection is used more.

Hit and false alarm rates are normally calculated as aggregate proportions, with the false alarm rate indicating the number of false positive responses out of all cases where the correct response is negative, and the hit rate indicating the number of correct identifications out of all instances where the correct response is positive. These are usually calculated for each individual and then aggregated across categories of interest such as experimental condition. In the analyses in this thesis changes in false alarms and hits were still used to index the extent to which participants used recollection, but rather than calculating an aggregate hit rate and false alarm rate, logistic mixed effects models were created with either false alarms or hits as the outcome. Hence analyses were based on the individual data points rather than aggregate proportions, avoiding the problems with use of aggregate statistics discussed earlier. In the first set of experiments, false alarm and hit rates and *d* prime values are presented for comparison with earlier research. However, since the first experiments demonstrate that both methods lead to the same conclusions this in not continued throughout the thesis.

In summary, data analysis throughout this thesis used a logistic mixed effects modelling approach, due to the risks of using more traditional approaches for categorical outcome variables, and the benefits of controlling for variation between stimuli and individual participants. Some data on false alarm and hit rates will be presented for the first set of data to demonstrate consistency with previous findings for the types of tasks used.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Bates, D. M. (2008). Computational methods for mixed models [Tech. Rep.]. Madison, WI: Dept. of Statistics, University of Wisconsin.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of experimental psychology. Applied, 8*(1), 44-56.

Canty, A., & Ripley, B. (2014). boot: Bootstrap R (S-Plus) Functions.

Canty, Angelo, & Ripley, Brian. (2013). boot: Botstrap R (S-Plus) functions. Retrieved from http://cran.r-project.org/

Davison, A. C., & Hinkley, D. V. . (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology-Learning Memory and Cognition, 35*(1), 57-80. doi: 10.1037/a0013865

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition, 36*(7), 1351-1359. doi: 10.3758/mc.36.7.1351

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*(4), 705-709. doi: 10.3758/bf03208849

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi: 10.1016/j.jml.2007.11.007

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of experimental psychology. Learning, memory, and cognition, 22*(5), 1304-1316.

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology-Learning Memory and Cognition, 40*(5), 1287-1306. doi: 10.1037/a0036914

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of applied psychology, 85*(4), 504-511.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of*

Experimental Psychology-Learning Memory and Cognition, 26(2), 294-320. doi: 10.1037/0278-7393.26.2.294

Team, R Core. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing

Retrieved from http://www.R-project.org.

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35*(3), 526-544. doi: 10.3758/bf03193292

CHAPTER 3 — COMPARING ITEM AND ASSOCIATIVE TASKS

This thesis is designed to test the theory that the positive-negative difference is caused by neglect of recollected evidence in recognition decisions. This theory prompts a basic prediction which I aimed to test in the first experiments. As I proposed that the positive-negative difference is caused by an underlying process involved in all recognition decisions, it should occur in all recognition tasks to different degrees depending on the extent to which recollected evidence is available and/or used in those tasks. There are two potential influences outlined in the literature which may regulate the extent to which recollected evidence is used. Firstly, the availability of cues to assist with recollection varies between tasks. Item recognition tasks, a list of items are studied one at a time without any accompanying information. If at test a studied item is presented, it provides a cue to enable recollection of information indicating that it was studied, but also produces familiarity which may be adequate for a decision. When an unstudied item is presented at test, however it provides no cue to recollected information, since it is not relevant to any of the studied items. On the other hand, in associative tasks, where items are paired together or paired with contextual information such as the modality of presentation, details are provided at test which can act as cues for recollection. For example, in a word pair recognition task, an individual may recollect imagining the two items together when they studied two nouns. Therefore, when one item is presented at test, it can act as a cue to help the individual recollect the other item it was paired with. This allows both matching and mismatching between the pair item for which information is recollected and the suggested pair item shown at test. Therefore confidence in negative decisions can be based on mismatch, rather than an absence of evidence for a positive decision.

Secondly, individuals moderate their use of recollection depending on what is most efficient for the task (Malmberg, 2008; Malmberg & Xu, 2007). In item recognition tasks where items are either studied or unstudied at test, little recollected information is required for the task to be completed successfully, as the relative familiarity of items provides evidence of which ones have been recently studied. In contrast , in associative tasks recollected information is important for distinguishing similar studied foils from studied stimuli, as similar foils are also familiar. Individuals may choose not to use recollected evidence in situations where they perceive it to be less efficient or necessary for the task, and therefore recollected evidence is likely to be neglected in item recognition.

Neglect of recollection can be expected to lead to changes in the confidence-accuracy relationship. Where recollection is used, confidence can be based on the degree of match or mismatch between the recollected detail and the test item. This provides strong evidence for the decision made, and for the confidence rating given. However, if recollected evidence is not used in a decision confidence cannot be based on the degree of match or mismatch, and can only be based on the amount of familiarity associated with the test item. This leads to negative decisions being based on lack of evidence for a positive decision rather than a mismatch with memory. Therefore, tasks which involve more recollection could be expected to show a stronger confidence-accuracy relationship, particularly for negative decisions.

The role of my first study was to examine the confidence-accuracy relationship in item recognition for words. A basic prediction of my theory is that the positive-negative difference should be evident in a variety of recognition tasks when recollected evidence is neglected, not only in face recognition. Therefore any item recognition task should show a positive-negative difference regardless of the type of stimuli. This study tested this prediction. It was important to establish whether the positive-negative difference occurred for words as stimuli as later manipulations of recollection I was hoping to use would be easier to implement with words. I expected that the confidence-accuracy and response-latency-accuracy relationships in the word item recognition task would be stronger for positive decisions than negative decisions, therefore showing a positive-negative difference. Since item recognition is not expected to utilise recollected evidence for the reasons discussed above, this task was expected to provide an example of the confidence-accuracy relationship in a task where little or no recollection is used. If the positive-negative difference is due to neglect of recollected evidence in recognition decisions, it should appear strongly in item recognition tasks where recollected information used is at a minimum.

Study 1: Item recognition

Method

Participants. Participants were 40 Flinders University students (32 female), ranging in age from 17 to 41 years, who participated in return for either course credit or remuneration. Participants were required to have English as a first language and normal or corrected-to-normal vision.

Materials. Stimuli were 128 English nouns between 4 and 10 letters in length, with similar Kucera-Francis written frequencies (250-400) concreteness (300-500) and imagability (300-500) ratings, sourced from the University of Western Australia’s MRC Psycholinguistic Database (Wilson, 1988). Words were randomly allocated by the experimental software to be studied or unstudied for each participant.

Procedure. All participants gave informed consent before participating after reading a brief description of the types of tasks involved. Participants completed the entire task on computers in individual cubicles. During the study phase participants were presented with 64 of the nouns which appeared one at a time in the centre of the computer screen for 1000 ms with a 1000 ms inter-stimulus interval during which the screen was blank. In the test phase, participants completed 64 trials. In each trial a word appeared in the centre of the screen with *yes* and *no* answer buttons displayed below. Participants were told to select *yes* if the word on screen was one they had studied, and *no* if the word was not on the study list. When participants clicked on a response button their response latency was recorded by the computer (timing started when the screen with the word and response buttons first appeared, and stopped when a response button was clicked). A confidence scale then appeared on the screen for participants to rate their certainty that their answer was correct. Participants were given instructions for how to use the confidence scale at the beginning of the test phase. They were told that they should select the rating that best indicated their confidence that the answer they just gave was correct, with 50% representing a guess and 100% representing certainty that they were right. Confidence rating buttons were presented in 10% intervals from 50% to 100%.

Results

To test the hypothesis that a positive-negative difference would be present in the confidence-accuracy relationship, a mixed effects model was created with confidence and response type (positive or negative) as a fixed effects and accuracy as the outcome, following the procedure outlined in Chapter 2. The fixed effects coefficients for this model are presented in Table 2. A positive-negative difference would be shown by an interaction between confidence and response type, since confidence would be expected to predict accuracy for positive decisions, but either not predict or less strongly predict accuracy for negative decisions. Although confidence buttons ranged from 50 to 100 in intervals of 10, the confidence values recorded in the dataset ranged from 0 to 5 with 0 representing a rating of 50% and 5 a rating of 100%. The estimate of -0.39 shown in the table indicates that for each increasing 10% unit of confidence, the predicted log odds of an accurate decision increased

.388 less for negative decisions than for positive decisions. This is depicted in Figure 1. The confidence interval for this interaction [-0.51, -0.27] does not include zero, indicating a

Table 2

Fixed Effects Coefficients for Confidence [a] *and Accuracy and Response Latency and Accuracy in Study1 by Response Type*

Fixed effect	*b*	SE_b	95% CI
	Confidence and accuracy		
Intercept	-0.74	0.16	[-0.98, -0.53]
Response: Negative (R)	1.34	0.20	[0.97, 1.74]
Confidence (C)	0.52	0.05	[0.44, 0.61]
R × C	-0.39	0.06	[-0.51, -0.27]
	Response latency and accuracy		
Intercept	0.39	0.20	[0.09, 0.68]
R	0.002	0.17	[--0.31, 0.34]
Response latency (L)	-0.08	0.05	[-0.18, 0.02]
R × L	0.08	0.07	[-0.07, 0.21]

Note. The models included participant (*SD* = 0.17) and question (*SD* = 0.32) as random effects.

significant effect. Hence confidence predicted accuracy more strongly for positive than negative decisions, as predicted. Therefore the hypothesis that a positive-negative difference for confidence would be seen in this task is supported.

A mixed effects model was also created to examine whether a positive-negative difference also occurred in the response latency-accuracy relationship. The fixed effects coefficients for this analysis are also shown in Table 2. Response type (positive or negative) and response latency (converted to *s*, rounded to 3 decimal places) were added as fixed

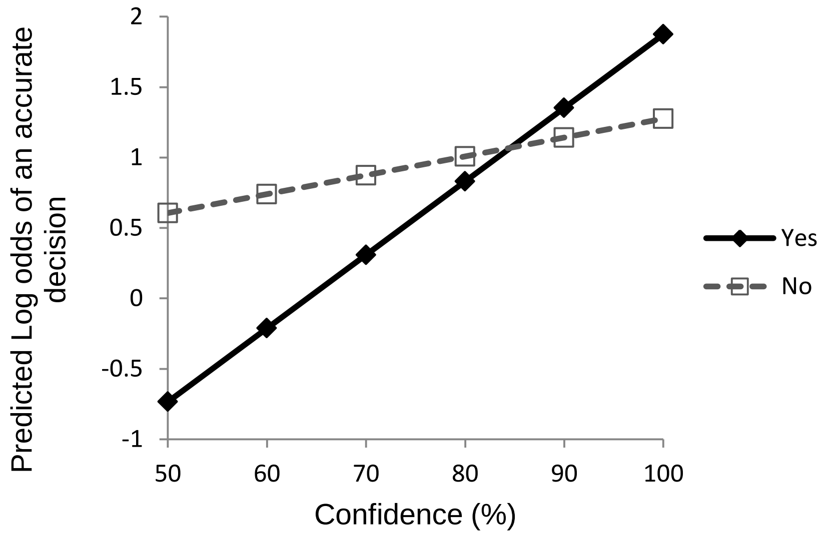

Figure 1. Predicted log odds of an accurate decision by confidence, for positive and negative decisions separately.

effects with accuracy as the outcome. Here the estimate of change in accuracy with each *s* increase in response latency for baseline (positive) decisions is -0.08 (95% *CI* = -0.18, -0.02). The negative value of this estimate shows that more rapid decisions were more likely to be accurate, consistent with previous research, however as the confidence interval includes zero this effect was not significant. The estimate of change in this relationship for negative, compared with positive decisions is shown by the interaction of 0.08 (95% CI = -0.07, 0.21). This indicates that for each *s* change in response latency the predicted log odds of an accurate decision decreased by .08 less for negative than for positive decisions. This demonstrates that response latency was a weaker predictor of accuracy for negative decisions, consistent with a positive-negative difference. However, as the 95% confidence interval includes zero, this effect also was not significant.

Discussion

The results of Study1 indicate that the positive-negative difference does generalise to words. This supports the idea that it may be caused by an underlying mechanism involved in all types of recognition. Faces and words as stimuli differ in a number of ways and are more different from one another than faces are from other pictorial stimuli which share more identifying features such as shape and colouring. Therefore the finding of a positive-negative difference in word recognition supports the idea that it stems from common processes involved in all recognition decisions. If this is the case, the effect could be expected to generalise to other kinds of stimuli also.

The fact that the positive-negative difference occurs in an item recognition task, in which very little recollected information is available to individuals, is consistent with the idea that neglect of recollected evidence causes the poor confidence-accuracy relationship for negative decisions. According to my theory, the lack of recollected evidence available in item recognition leads decisions to be based on familiarity. Therefore decisions and their subsequent confidence ratings can only be based on the amount of evidence or lack of evidence for a positive decision. Hence confidence in negative decisions is based on lack of evidence. Because of this I would expect confidence to be less strongly related to accuracy for negative decisions than positive decisions. Therefore the fact that confidence and accuracy were less strongly related for negative decisions in the word item recognition task is consistent with my explanation.

The interaction between response and confidence for response latency was also in the direction of a weaker relationship for negative decisions, which is consistent with the expectation that response latency will follow similar patterns to confidence. However, this effect was not significant, with the lower estimate of the confidence interval falling just below zero. This suggests that the effect may be very small compared with that for confidence, or may have occurred by chance. Hence the response-latency-accuracy relationship did appear to show similar patterns to the confidence-accuracy relationship in this task, but conclusions are limited, and a detailed explanation of response latency is beyond the scope of this thesis. Predictions concerning the response-latency accuracy relationship are not straightforward, since in associative tasks, where recollection is more crucial to accuracy, response latency is likely to increase for decisions where recollection is used due to the process being slower, as discussed in the general introduction. Hence it is difficult to study the response-latency-accuracy relationship in an item recognition task alone.

This study demonstrated that the positive-negative difference in the confidence-accuracy relationship generalises to words, which supports the idea that it is caused by basic mechanisms underlying recognition decisions. However in order to test the idea that the amount of recollection used influences the degree to which the positive-negative difference occurs, it was necessary to investigate the positive-negative difference in a task that allows use of recollected evidence. An example of this is an associative recognition task, in which foil and target stimuli are difficult to distinguish without use of recollection.

Experiments 2a and 2b: Plurality discrimination

From Study1 I was able to show that the positive-negative difference occurs in item recognition using words, where the amount of recollected evidence available is minimal. Associative recognition, on the other hand, has more recollected information available and requires individuals to use recollection to achieve accuracy. Therefore comparing the results of Study1 with the confidence-accuracy relationship seen in an associative task was expected to demonstrate a smaller positive-negative difference in the associative task, due to increased use of recollection. The aim of Experiments 2a and 2b was to provide such an associative task.

Plurality discrimination is a type of associative recognition task where recollection is particularly important to accuracy, as targets and foils are very similar and difficult to discriminate on the basis of familiarity. In a plurality discrimination task participants study words in either singular or plural form and at test are asked to only give a positive response to words that are the same plurality as they were studied. Words which are presented in the test phase with changed plurality, for example ‘windows’ instead of ‘window’ act as foils. Because only the presence of absence of the letter ‘s’ varies between targets and foils, it is very difficult to complete this task without recollection, since the opposite plurality form of a word is likely to also become familiar when the word is studied. In Experiments 2a and 2b a plurality discrimination task was used as an associative task in which participants were expected to use at least some recollection. This was expected because a) recollected information would be more available (participants would have a cue for retrieval of recollected information provided by either the studied word itself or a very similar word), and b) as suggested by Malmberg (2008) participants would be likely to perceive that the task would be difficult without recollection, and therefore the most efficient way to complete the task would be to use recollected information at test.

To determine whether recollected evidence was used in the associative recognition decisions, I also included a manipulation of familiarity. When an item at test is unstudied, but very similar to a studied item, familiarity is still high, especially when familiarity of the similar item studied is increased. Recollected evidence provides the mismatching details between the highly similar, and therefore familiar foil, which allows individuals to avoid false alarms to these items. Therefore, when recollection is used, the effects can be seen in the types of decisions made about these new but similar stimuli. Where recollected information used less, increasing the familiarity of similar items will lead to increasing false alarms to the new but similar items at test. As the amount of recollected detail used in decisions increases, false alarms to these similar items should no longer be affected by increasing familiarity. Hence recollection can be expected to counter the effect of increasing familiarity of similar, but unstudied, stimuli (such as the plurality-altered versions of studied words) by providing recollected detail that informs the individual that the similar stimulus is not the same as the studied stimulus. For example, if an individual recollects studying the word *tree* they can avoid making a false positive response to the word *trees* even though *trees* may become familiar when *tree* is studied. If the familiarity of *tree* is manipulated, this should not change the likelihood that participants will respond *yes* to *trees*. If, on the other hand, an individual is not using enough recollected evidence, and instead relies on familiarity, the more familiar the word *tree* is made to be, the more familiar *trees* is also likely to become. Therefore, as the familiarity of *tree* is increased, false alarms to the word *trees* will also increase. This means that by manipulating familiarity, it is possible to determine whether recollected evidence is involved by observing the pattern of false alarms.

Repetition of a stimulus is a manipulation that is frequently used in recognition memory experiments to increase the strength or familiarity of an item (e.g., Malmberg, Holden, & Shiffrin, 2004; Singer, Fazaluddin, & Andrew, 2012). Its effects can vary depending on the extent to which participants are using recollection compared with familiarity (Hintzman, Curran, & Oppy, 1992; Jacoby, Jones, & Dolan, 1998). Although in my experiments strengthening items through repetition was used to highlight the effects of increased familiarity, repetition increases memory strength overall, and hence can also affect recollection (Jones & Jacoby, 2001). If individuals are using recollected evidence in their decisions, increased opportunities to encode a stimulus should lead to improved ability to recollect that stimulus, and hence false alarms should be unaffected or decrease while hit rates increase. If, however, individuals are relying on familiarity, increased presentations of a stimulus can increase the familiarity of all similar stimuli. Therefore if recollection is not used, and the task involves distinguishing similar items, individuals are likely to make more false alarms with increasing presentations of similar stimuli.

Plurality discrimination is a task in which recollected evidence is available, as the test items share features with studied items even when they are not targets, which allows them to act as cues. In addition, plurality discrimination is difficult to complete without recollection, as foil items are highly similar to targets, as with most associative recognition tasks (Hockley & Consoli, 1999; Malmberg, 2008; Rotello & Heit, 2000; Rotello, Macmillan, & Van Tassel, 2000). Therefore I expected participants to use recollection in this task. Three possible patterns of results could occur. I predicted that recollection would occur in the plurality discrimination task. This could be reflected in two of the possible patterns: First, with high use of recollection, increasing presentations of stimuli would lead to increased ability to recollect the presented stimulus and hence increased hits and reduced false alarms. Second, with intermediate levels of recollection, false alarms may neither significantly increase or decrease with increasing presentations. This would still support my prediction that participants use recollection in this task. However, if false alarms increased with increasing presentations, this would demonstrate that participants were using very little recollection in this task.

I hypothesised that increased use of recollection in the plurality discrimination task would result in a reduced or non-significant positive-negative difference. Since my theory suggests that the positive-negative difference results from neglect of recollected evidence, a task which results in more use of recollected information should also result in a stronger confidence-accuracy relationship for negative decisions, and therefore a reduced positive-negative difference. Two small experiments (2a and 2b) were conducted which differed only in the respect that one had more specific instructions before the study phase advising participants that the plurality of the words would be important at test. The purpose of the more specific instructions was to investigate whether participants' awareness that memory of plurality was important would influence the extent to which they used recollected evidence. However, the pattern of results in the two versions of the experiment were the same, so the main analyses presented are for the data from both experiments pooled.

Method

Participants. Nineteen undergraduate Flinders University students (13 female) participated in experiment 2a, and 21 (14 female) in experiment 2b. Participants ranged in age from 17 to 45 years (M = 22.78, SD = 6.33). All were required to have English as a first language and normal or corrected-to-normal vision.

Materials. Stimuli were 128 nouns and their plural forms, taken from the MRC Psycholinguistic Database, balanced for written frequency, imagability and concreteness. For each participant, the word list was randomly assigned into 4 blocks. Within each block, words were randomly assigned to be presented in the studied form at test, one quarter to be presented with changed plurality, and one quarter to be shown at test as completely new words. Each word allocated to be studied was then randomly assigned to be presented in plural or singular form, and to be presented once or three times.

Procedure. Experiments 2a and 2b were the same except that in the second version of this experiment, the pre-study instructions to participants were altered to draw their attention to the fact that the plurality of the words would be important in the memory test, by adding the instruction: *Try to note whether each word is singular or plural. The plurality of the words will be important when your memory for these words is tested later in the session*.

Each experiment was broken into four study and test phase blocks. In each study block participants saw 24 words, which appeared one at a time on the computer screen, for a duration of 1 *s* per presentation. In the test phase, Participants were asked to respond *yes* only to words that were exactly as studied, and to respond *no* to words that had changed plurality or were new. During each test trial, a word appeared on the screen, with response buttons displayed below. Once participants clicked on a response button, they were asked to indicate their confidence that their decision was correct by clicking on a confidence button in 10% intervals from 50-100%. The time taken for participants to respond to both the word and the confidence scale was recorded by the computer program, with timing starting when the screen showing the response buttons appeared and stopping when participants clicked on a response button.

Results

Data from the two experiments were combined for analysis due to the minimal difference between the two experiments. To test whether the experiments had significantly different results, all analyses were initially run with experiment as a fixed effect. Interactions between experiment and the other predictor variables were also considered, including other interaction terms. For most analyses the effect of experiment was not significant, and none of the interactions were significant, and therefore these results are not presented here. For those analyses where there was a significant effect of experiment this will be noted in the results. Full results for the analyses including experiment are presented in Appendix 3A.

To test the degree to which participants were using recollection in these experiments, two mixed effects models were created with number of presentations entered as a fixed effect. The first model had false alarms as the outcome and the second hits as the outcome. If participants were not using recollection at all, I would expect to see false alarms and hits increasing equally with number of presentations. If participants used some recollected information, I would expect hits to increase more than false alarms as presentations increased, and if participants used the maximum possible amount of recollection, I would expect only hits to increase with presentations and false alarms could potentially decrease with presentations as more recollected information would become available.

Table 3 shows the fixed effects coefficients for the models predicting false alarms and hits. Table 4 presents the proportion false alarm and hit rates in the different presentation conditions for comparison with earlier research. The predicted log odds of a false alarm occurring increased significantly with number of presentations as did the predicted log odds of a hit. The *b* value of 1.69 (95% CI = 1.45, 1.93) for the change in hits with three presentations demonstrates a larger change in hits than the comparative change in false alarms (*b* = 0.34, 95% CI = 0.09, 0.59), as a) the confidence intervals do not overlap and b) the lower estimate for the change in false alarms is close to zero. The comparatively large increase in hits suggests that some recollected evidence may have been used, although it was not enough to eliminate an increase in false alarms[6], participants used some recollection in this task.

To examine the confidence-accuracy relationship for negative decisions in the plurality discrimination task, I created a mixed effects model was created with confidence and

[6] One explanation of a greater change in the hit rate is that participants simply shifted their response criterion downward in a distribution where the signal and noise familiarity do not overlap to a large degree. This would result in most of the old items being given positive responses with very little change in the number of false alarms. However this is not plausible for the data observed, due to the size of the false alarm rate (.30 [SD = 19], changing to .36 [SD= .21] following three presentations) showing that false alarms were not at a minimum rate (that is, not values from the tail of the noise distribution), and therefore movement in the criterion would be expected to increase false alarms more substantially compared with hits if recollection evidence were not being used.

Table 3

Fixed Effect Coefficients for Logistic Mixed-Effects Models Predicting false alarms and hits by presentations in Experiments 2a and 2b combined

Fixed effect	*b*	SE_b	95% CI
	False alarms		
Intercept	-0.98	0.16	[-1.30, -0.67]
Presentations	0.34	0.15	0.09, 0.59
	Hits		
Intercept	0.95	0.13	[0.70, 1.21]
Presentations	1.69	0.12	1.45, 1.93

Note. The models included participant (*SD* = 0.81) and test word (*SD* = 0.27) as random slopes and intercepts.

Table 4

False alarm rates, hit rates and d prime (standard deviations in brackets) in experiments 2a and 2b combined by number of presentations

	FA rate	Hit Rate	*d'*
1 presentation	.30 (.19)	.70 (.15)	1.19 (0.82)
3 presentations	.36 (.21)	.92 (.07)	1.87 (0.83)

response as predictors and accuracy as the outcome (see Table 5 for fixed effects coefficients).Figure 2 compares positive and negative decisions in the item recognition task (dotted lines) with positive and negative decisions in the plurality discrimination task (solid lines). The figures show the change in the positive-negative difference between the item recognition task, where very little recollected evidence was available, and the associative task, where recollected evidence was available. Although the item recognition and plurality

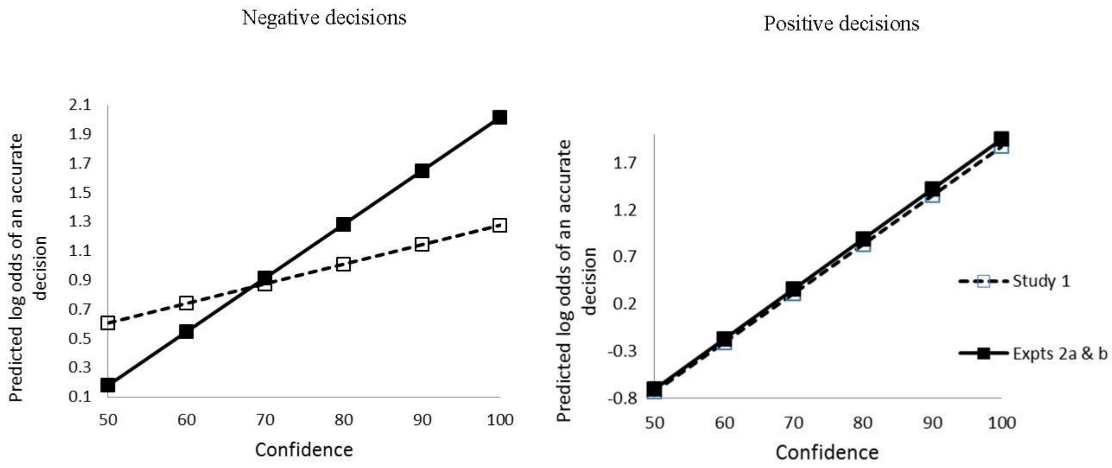

Figure 2. Predicted log odds of an accurate decision by confidence in the item recognition task using words (Study 1, shown by dotted lines) and the plurality discrimination task (Expts 2a and 2b, shown by the solid lines), separately for positive and negative decisions.

discrimination procedures are not directly comparable, the difference suggests that participants' confidence in negative decisions was more strongly related to accuracy when partial recollection was used than when no recollection was used. These results provide some support for the idea that increased use of recollected evidence improves the confidence-accuracy relationship for negative decisions, decreasing the positive-negative difference. A similar mixed effects model was created with response latency in the place of confidence (also shown in Table 5), to investigate whether the response latency-accuracy relationship changed in similar ways to that for confidence and accuracy. The 95% confidence interval for the interaction between response latency and response includes zero (-0.079, 0.076). Therefore there was no significant effect of decision type (positive versus negative) on the response-latency-accuracy relationship. The very small *b* value and the fact that both bounds of the confidence interval are close to zero suggests that if there was an effect of decision

Table 5

Fixed effects coefficients for mixed effects models predicting accuracy from confidence and response and response latency and response.

Fixed Effect	*b*	SE_b	95% CI
	Confidence		
Intercept	-0.91	0.18	[-1.37, -0.50]
Confidence (C)	0.59	0.05	[0.46, 0.60]
Response (negative) (R)	0.99	0.19	[0.54, 1.23]
C × R	-0.19	0.05	[-0.25, -0.08]
	Response latency		
Intercept	-1.06	0.25	[-1.39, -0.54]
Response latency (RL)	-0.18	0.03	[-0.24, -0.12]
R	0.11	0.12	[-0.12, 0.34]
RL × R	-0.002	0.04	[-0.08, 0.08]

Note. The models included participant (*SD* = 0.53) and test word (*SD* < 0.01) as random slopes and intercepts.

type on the response latency-accuracy relationship, it was very small. The negative value of *b* suggests that the response latency-accuracy relationship may have been weaker for negative decisions, which suggests a possible trend towards a pattern similar to the confidence-accuracy relationship. However, since the effect was not significant no clear conclusions can be drawn regarding the similarity of the response latency-accuracy and confidence-accuracy relationships.

General Discussion

Overall the data from these experiments demonstrated that the positive-negative difference can occur in word recognition tasks, and in both item and associative recognition. A positive-negative difference in item recognition was expected due to the lack of recollected evidence available for this task and the relative ease with which it may be completed using familiarity. However, in associative recognition, participants could be expected to use more recollection since a) more recollected evidence is available, and b) it is difficult to complete the task accurately without using recollection. The positive-negative difference observed in the associative recognition task was smaller, consistent with increased use of recollection. However the fact that false alarms still increased with presentations suggests that recollection was not used in all decisions. The fact that a positive-negative difference was present, albeit smaller, is consistent with the theory being tested given that the pattern of false alarms suggests recollection was only partially used. There are two possible reasons for only partial use of recollection in the associative task. Firstly, recollected information may not be equally available for all decisions. Targets presented at test may provide a stronger cue for recollection than do foils, since these stimuli are full matches for items in memory, making use of recollected evidence in positive decisions easier than use of the same evidence in negative decisions. Secondly, participants may not perceive recollected evidence as necessary for all decisions, or they may choose to make a decision before recollected evidence has accumulated in order to respond in an efficient manner. If this latter suggestion underlies participants' neglect of recollected evidence in associative tasks, manipulations which encourage use of recollection may improve the confidence-accuracy relationship for negative decisions. However, in these experiments it was not possible to directly observe the degree to which recollection was used and the effects of this on the confidence-accuracy relationship. In order to study this relationship directly, manipulation of availability of recollection would be necessary. These patterns of false alarms do however support the idea that partial recollection occurred in the associative task, in contrast with the item recognition task.

The difference in the confidence-accuracy relationship between Study1 and Experiments 2a and 2b demonstrates the difference between item recognition, where recollection is unlikely to be used at all, and the patterns of false alarms suggest it was not, and the associative recognition task, where some recollection is necessary for any level of accuracy, and false alarms suggested partial recollection was used. As Figure 2 depicts, this difference was mainly in the confidence-accuracy relationship for negative decisions, with a stronger relationship between confidence and accuracy occurring for negative decisions in the plurality discrimination task. This is consistent with increases in use of recollected evidence resulting in an improved confidence-accuracy relationship for negative decisions.

While this provides some preliminary support for the theory that the positive-negative difference is produced by neglect of recollection, the results do not provide clear evidence that the difference between the two tasks was due to differences in recollection, and not to other features that differ between item and associative recognition tasks. In order to provide such evidence, manipulation of recollection within a single task was needed. This was the role of subsequent experiments.

References

Canty, Angelo, & Ripley, Brian. (2013). boot: Botstrap R (S-Plus) functions. Retrieved from http://cran.r-project.org/

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory - registration without learning. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(4), 667-680. doi: 10.1037/0278-7393.18.4.667

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27*(4), 657-664. doi: 10.3758/bf03211559

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*(4), 705-709. doi: 10.3758/bf03208849

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*(1), 82-102. doi: 10.1006/jmla.2000.2761

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology-Learning Memory and Cognition, 30*(2), 319-331. doi: 10.1037/0278-7393.30.2.319

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*(3), 545-556. doi: 10.3758/bf03193293

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*(6), 907-922. doi: 10.3758/bf03209339

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43*(1), 67-88. doi: 10.1006/jmla.1999.2701

Singer, M., Fazaluddin, A., & Andrew, K. N. (2012). Recognition of categorised words: Repetition effects in rote study. *Memory, 21*(4), 467-481. doi: 10.1080/09658211.2012.739625

Appendix 3A: Tables Showing Fixed Effects Coefficients for Analyses Presented in Chapter 3 Results With Experiment Included as a Fixed Effect

Table A1.

Fixed Effects Coefficients for Fitting Logistic Mixed-Effects Models Predicting False Alarms and Hits from Number of Presentations and Experiment

Variable	Estimate	SE	95% CI
	False alarms		
Presentations (P)	0.273	0.184	-0.0871, 0.6434
Experiment (E)	-0.197	0.316	-0.8251, 0.4323
P × E	0.124	0.256	-0.4032, 0.6406
	Hits		
P	1.642	0.186	1.220, 2.064
E	-0.315	0.245	-0.8425, 0.0580
P × E	0.079	0.247	-0.6213, 0.8125

Table A2.

Fixed Effects Coefficients for Fitting Logistic Mixed-Effects Models Predicting Accuracy from Confidence and Response Latency with Response, Deadline Condition and Experiment.

Variable	Estimate	SE	95% CI
	Confidence		
Confidence (C)	0.595	0.053	0.4405, 0.7048
Response (R)	1.388	0.276	0.648, 2.046
Experiment (E)	0.620	0.338	-0.0861, 1.0814
C × R	-0.246	0.070	-0.4017, -0.0790
R × E	-0.871	0.365	-1.447, -0.382
C× E	-0.105	0.072	-0.2480, 0.0445
C × R × E	0.135	0.094	-0.0497, 0.3445
	Response latency		
Response latency (RL)	-0.099	0.043	-0.1538, -0.0178
Response	0.458	0.168	0.1482, 0.7783
Experiment	0.464	0.262	0.0011, 0.7202
RL × R	-0.117	0.062	-0.2172, -0.0128
RL × E	-0.138	0.058	-0.2524, -0.0505
R × E	-0.651	0.231	-1.0187, -0.2330
RL × R× E	0.195	0.079	0.0643, 0.3334

CHAPTER 4 — RESTRICTING RECOLLECTION: PLURALITY DISCRIMINATION

The first set of experiments tested the confidence-accuracy relationship in an item recognition task, where little recollected information would be available or likely to be utilised by participants, compared with an associative recognition task, where recollected information was available and important for accuracy. The associative task was expected to result in greater use of recollection than the item recognition task for two reasons. First, information in the test stimulus in an associative task provides a cue for recollection in negative decisions as well as positive, since foils as well as targets have some features which match the studied stimulus. Second, as suggested by Malmberg (2008), individuals are more likely to use recollected evidence when they perceive it as more efficient. Since associative tasks are difficult to complete accurately without recollected evidence, individuals should perceive use of recollection as important to completing the task with an acceptable level of accuracy, and therefore more efficient in the context. Therefore individuals are more likely to utilise recollection in associative tasks.

I expected that reduced use of recollected evidence would result in a weaker confidence-accuracy relationship for negative decisions. Recollected evidence allows confidence in negative decisions to be based on evidence of a mismatch, so more evidence leads to higher confidence, as explained in previous chapters. However, where recollected evidence is not used, negative decisions based on lack of familiarity (lack of evidence for a positive decision) are more confident when less evidence is available. The comparison of the item recognition and associative recognition tasks provided some evidence for the theory that neglect of recollection causes the poor confidence-accuracy relationship for negative decisions. The confidence-accuracy relationship for negative decisions was stronger in the associative task, where patterns of false alarms suggested more recollection was used. However, since recollection was not manipulated in these experiments, it was not possible to clearly compare decisions in which participants used recollection, with those in which they did not, separately from other factors which may differ between the selected tasks. Therefore, a casual effect on the confidence-accuracy relationship could not be inferred. Experiments 3 and 4 were designed to tackle this issue by decreasing the availability of recollected evidence within an associative task. Thus, I could directly observe the impact on the relationship between confidence and accuracy. An associative task (plurality discrimination) was used because a) recollection can be expected to be used to some degree in these types of task, as explained earlier, and b) Experiments 2a and b demonstrated that some recollection is used in this task. Therefore, the amount of recollection available could be reduced without needing to first establish that participants could use recollected evidence when intended.

In order to specifically observe the effects of decreasing recollection on the confidence-accuracy relationship it was necessary to find a manipulation which would reduce recollection without reducing familiarity. To do this I turned to the literature on how recollection and familiarity differ in search of a variable which would affect recollection more than familiarity. One respect in which recollection and familiarity are often proposed to differ is the time course within which each type of evidence becomes available. For example, Hintzman and Curran (1994) demonstrated that participants were able to discriminate completely new words from studied words earlier than they could distinguish studied words from plurality-altered foils. New words share fewer features with studied stimuli, and therefore would be expected to be meaningfully less familiar than studied words. Therefore familiarity evidence is adequate to distinguish new words from studied words. On the other hand, plurality-altered foils share most of their features with studied items. This means that words which have changed plurality are highly similar in their familiarity to studied items. This makes them difficult to distinguish from studied words without use of recollection.

Therefore, Hintzman and Curran suggested that the delay in discriminating these similar foils from studied words was due to recollected evidence accumulating more slowly than familiarity evidence. In a variation on the 1994 study Hintzman, Caulton, and Levitin (1998) replicated this result. Participants were asked to distinguish between words studied on different study lists, as well as studied and completely new words. Words which have been studied, but on a different list, should have similar familiarity to the target words studied on a specific list. Therefore, they should be hard to distinguish without use of recollection. In contrast, completely new words which have not been recently studied can be expected to have meaningfully lower familiarity than words which have just been studied. Therefore completely new words should be distinguishable from words studied on the recent lists without use of recollection. Participants were found to take longer to discriminate between words studied on different lists than they did to discriminate old from entirely new words. This suggests that it took longer for participants to make the distinction when recollected evidence was needed.

Further evidence that recollection is slower to utilise than familiarity comes from a study by Light, Patterson, Chung, and Healy (2004). In a word pair study, (Light et al., 2004) asked participants to distinguish between intact and rearranged word pairs. Because both words in all word pairs tested were studied, all tested word pairs should be similarly familiar. Therefore this task could be expected to require recollection. Young adults were found to make more false alarms to re-arranged pairs under a short response deadline than under a long response deadline. This again suggests that these similar stimuli are difficult to distinguish rapidly, providing support for the idea that recollection accumulates later.

Another way that the longer onset of recollected evidence has been shown is in the patterns of false alarms to similar items. When familiarity is used similar foils like a reversed plurality word should seem highly familiar due to sharing the majority of their features with a studied item. This normally results in participants giving false positive responses to these similar items when they are unable to use recollection. In Light et.al.'s (2004), first experiment, longer response deadlines led to fewer false alarms to similar foils than shorter response deadlines. This suggests that increasing the time spent on a decision may make distinction between similar items easier, which is potentially due to recollected evidence being available. Similarly in a study by Jacoby, Jones, and Dolan (1998) participants were asked to distinguish words studied auditorially from those studied visually. Hence all words were expected to be familiar due to having been studied, and recollection was expected to be required to determine which words had been studied in a certain form. Participants made more false alarms to words which had been presented in the opposite form when under a short response deadline than when under a long response deadline. Hence evidence suggests that recollection is slower to accumulate than familiarity, and therefore decisions made under deadline pressure are less likely to use recollected evidence. Therefore, in Experiments 3 and 4 I used a response deadline to reduce recollection.

Experiments 3 and 4 used a response deadline to reduce the extent to which participants could use recollected evidence in a plurality discrimination task. Participants were expected to have less access to recollected information in the deadline condition, and increased access to recollection information in the delay condition, due to recollected evidence taking longer to accumulate. I expected that confidence and accuracy would be less strongly related for negative decisions when recollection was restricted, than when it was unrestricted. Therefore, considering only negative decisions, I predicted an interaction between confidence and deadline condition such that confidence is a stronger predictor of accuracy in the delay condition.

In addition, familiarity was manipulated through repeated presentations as in the previous associative experiments, in order to check whether recollection was changing between the conditions as expected. Words that were presented more times during study were expected to be more familiar, regardless of the form in which they were studied. Decreased use of recollection (expected under deadline conditions at test) would be reflected in increased false alarms to similar foils (words which had changed plurality) with increased presentations. These items were expected to increase in familiarity with increasing presentations of an item which shared some features with them (the reverse-plurality item). Increased use of recollected evidence, on the other hand (expected under delay conditions at test), could be expected to offset increasing familiarity, with recollection of the studied item becoming easier with increased presentations, and would be reflected by either no increase in false alarms with increasing presentations or a reduction in false alarms with increasing presentations.

Experiment 3

In Experiment 3 patterns of false alarms did not show any evidence of the response deadline manipulation restricting recollection. Therefore, only the method and manipulation check results are presented here, followed by Experiment 4 in which the manipulation was adjusted. Experiment 3 is underpowered due to the fact that I ended data collection early upon testing the manipulation, and finding that it was not effective in reducing recollection. I chose to adjust the manipulation and begin data collection for Experiment 4 rather than collect more data for a manipulation that was not successful.

Method

Participants. Thirty Flinders University students (26 female) ranging in age from 17 to 64 years (*M* = 22.96, *SD* = 10.40) participated in Experiment 3 in return for either course credit or remuneration. Requirements for English as a first language and vision were the same as for Experiment 2 (Chapter 3).

Materials. Stimuli were the same 128 nouns used in Experiment 2. As in Experiment 2, words were randomly allocated for each participant to be presented in singular or plural form, and to appear once or three times.

Procedure. Participants gave informed consent after reading a brief introduction describing the tasks involved. All participants completed the task on separate computers in individual cubicles. The experiment was in four study-test blocks, with the same study and test block pattern as for Experiment 2, except that at test, no new words were shown. At test, half of the words were presented as studied and half had changed plurality. Participants in the deadline condition were required to respond within 2 *s*. They were instructed that they would need to respond within this time for their answer to be recorded. If participants did not respond within 2 *s*, the recognition test screen disappeared and the confidence rating screen was displayed. Participants in the deadline condition were given the instruction that if this occurred, they should rate their confidence as 50%, as they had not given an answer. When this occurred the participant's response was coded as -1 and thus not included in the subsequent analyses. Participants in the delay condition were made to wait until after 2 *s* had passed to give their response. They were instructed that they would not be able to respond to each test word until after 2 *s* had passed and that they should wait until the response buttons appeared to give their answer. During the 2 *s* delay, the test item was displayed but the response buttons did not appear for participants to click on. After 2 *s* had passed, the response buttons appeared. Participants were told that they should try to respond as quickly as possible once the response buttons were present. Following their recognition decision, participants were asked to rate their confidence that their response was correct on a scale of 50-100%. Response buttons were displayed in 10% intervals. Participants were told in the instructions at the beginning of the test phase that they should choose 50% for a timeout, or if guessing, and 100% if they were certain they had selected the correct answer. The latency of both the recognition response and the confidence response were recorded, with timing starting when the test stimulus or confidence scale appeared on screen and stopping when the participant clicked on a response button.

Results

Manipulation checks. To test whether participants' ability to use recollection was impaired in the deadline condition, a mixed-effects model was created with false alarms as the outcome and number of presentations and condition as fixed effects. The fixed effects coefficients for this model are shown in Table 6 If the deadline manipulation impaired

Table 6

Fixed Effects Coefficients for Mixed Effects Models Predicting False Alarms and Hits from Condition and Number of Presentations in Experiment 3.

Fixed effect	*b*	SE_b	*95% CI*[a]
	False alarms		
Intercept	-0.50	0.10	[-0.78, -0.27]
Condition, delay (C)	-0.25	0.29	[-0.60, 0.19]
Presentations, 3 (P)	-0.06	0.15	[-0.33, 0.14]
C×P	0.16	0.21	[-0.17, 0.54]
	Hits		
Intercept	0.68	0.10	[0.41, 0.97]
C	-0.36	0.24	[-0.70, 0.10]
P	0.97	0.16	[0.51, 1.27]
C×P	-0.03	0.23	[-0.24, 0.32]

Note. The models included test word (*SD* = 0.30) as a random slope and intercept.

recollection,I expected to see a greater increase in false alarms with number of presentations in the deadline condition than in the delay condition. This would be shown by a negative *b* value for the interaction, which would indicate that false alarms increased less with presentations in the delay condition. The confidence interval for the interaction between presentations and condition includes zero, indicating that the delay and deadline conditions did not significantly differ in the increase in false alarms with presentations. In addition, the positive *b* value for the interaction indicates that if there was any effect, the model suggested it was in the opposite direction to what I expected. Therefore it could not be determined that the manipulation impaired recollection successfully, and confidence-accuracy data was not analysed.

Experiment 4

In Experiment 4, the familiarity manipulation was adjusted to ensure that all words had a higher level of familiarity. This was to ensure that participants could use familiarity to complete the task in the deadline condition, and prevent low memory quality from producing a floor effect. As in Experiment 3 the familiarity of half of the words was increased further, so that the effects of familiarity could still be distinguished to check whether the manipulation successfully reduced recollection. The procedure was the same as for Experiment 3, except that participants now studied the words three or six times.

Method

Participants. Eighty Flinders University students (61 female), ranging in age from 18 to 51 years (*M* = 21.86, *SD* = 5.77) participated in return for either course credit or remuneration. Again participants were required to have English as a first language and normal or corrected-to-normal vision.

Materials. Stimuli were the same as for Experiment 3, and were randomly allocated in the same way for each participant except that words were now allocated to be studied either three or six times.

Procedure. The procedure was the same as for Experiment 3 except for the additional presentations of words.

Results and Discussion.

One participant in the deadline condition gave no responses within the response deadline, and therefore provided no analysable data.

Manipulation checks. To test the degree to which recollection was impaired in the deadline condition and encouraged in the delay condition, a mixed effects model was created with false alarms as the outcome and condition and number of presentations as fixed effects. The fixed effects coefficients for this model and the same model created with hits as the outcome are displayed in Table 7. As with the data from Experiment 3, increased recollection in the delay condition was expected to be shown by a reduced or absent increase in false alarms with number of presentations, while in the deadline condition false alarms were expected to increase with presentations. This would be shown if a significant negative interaction occurred. The negative value of *b* for the interaction shows that as expected, false alarms in the delay condition increased less with increased presentations than in the deadline condition. The 95% confidence interval for the interaction does not contain zero, indicating that this effect was significant. For the model with hits as the outcome, the 95% confidence interval for the interaction contains zero, indicating that the change in hits as the number of presentations increased did not significantly differ between conditions. Table 8 shows a comparison of the hit and false alarm rates in each condition, and the associated *d* prime values, which support the conclusions drawn from the mixed effects model with the false alarm rate increasing with number of presentations only in the deadline condition.

Table 7

Fixed Effects Coefficients for the Mixed Effects Models Predicting False Alarms and Hits from Condition and Number of Presentations in Experiment 4

Fixed Effect	*b*	SE_b	*95% CI*
	False alarms		
Intercept	-0.50	0.13	[-0.81, -0.07]
Condition, delay (C)	-0.49	0.19	[-0.88, -0.08]
Presentations, 6 (P)	0.26	0.09	[0.09, 0.44]
C×P	-0.28	0.13	[-0.54, -0.07]
	Hits		
Intercept	0.68	0.11	[0.46, 0.80]
C	0.43	0.20	[0.05, 0.81]
P	1.04	0.11	[0.83, 1.25]
C×P	-0.04	0.15	[-0.35, 0.25]

Note. The models included test word (*SD* = 0.27) as a random slope and intercept.

To test whether the changes in false alarms were simply due to a change in response bias, a mixed effects model was created with response as the outcome and correct answer, condition and presentations as fixed effects. The fixed effects coefficients for this model are

Table 8

False Alarm Rates, Hit Rates and d prime by Condition and Number of Presentations in Experiment 4

Condition	Presentations	FA rate	Hit rate	*d'*
Deadline	3	.32(.15)	.67(.17)	1.01(.83)
	6	.41 (.20)	.78(.12)	1.15 (0.90)
Delay	3	.29(.13)	.72(.16)	1.28(0.79)
	6	.29(.18)	.86(.12)	1.85(0.94)

shown in Table 9. The regression coefficient for correct answer shows the change in the likelihood of a negative decision when the correct answer changed from being positive to being negative, in the deadline condition when the stimulus had been presented 3 times (coded as 0). The *b* value is positive, showing that a negative decision was more likely when it was the correct response. The 95% confidence interval does not include zero, hence this effect is significant. This shows that participants were able to discriminate studied and unstudied stimuli even when under pressure to make a rapid decision and limited by having studied the correct stimulus fewer times.

The *b* value for the fixed effect of condition shows the change in the predicted log odds of a participant making a negative decision between the deadline and delay conditions, when

Table 9

Fixed Effects Coefficients for the Mixed Effects Model Predicting Response from Correct Answer, Condition and Number of Presentations for Negative Decisions

Fixed Effect	*b*	SE_b	95% CI
Intercept	-0.62	0.06	[-0.71, -0.51]
Answer (A, negative)	1.08	0.09	[0.90, 1.22]
Condition (cond, delay)	-0.34	0.09	[-0.47, -0.23]
Presentations (6, P)	-0.97	0.10	[-1.15, -0.85]
A × Cond	0.75	0.12	[0.57, 0.93]
A × P	0.74	0.13	[0.57, 0.98]
Cond × P	0.08	0.14	[-0.09, 0.27]
A × Cond× P	0.17	0.19	[-0.14, 0.45]

Note. The model included test word (*SD* = 0.16) as a random slope and intercept.

the correct answer was positive (0), and a stimulus had been studied three times (0). This regression coefficient is negative, indicating that participants were less likely to make incorrect negative decisions in the delay condition, for stimuli that had been presented three times. As the 95% confidence interval does not include zero, this effect is significant.

For the fixed effect of presentations, the regression coefficient represents the change in the predicted log odds of a negative decision when presentations increased from three to six, in the deadline condition when the correct answer was positive. This *b* value is negative, indicating that when the stimulus had been presented six times, participants were less likely to make an incorrect negative decision. This effect is also significant, as the 95% confidence interval does not include zero.

The correct answer by condition interaction shows the change in the effect of the correct answer being negative when condition changed from deadline to delay. The *b* value is positive indicating that the predicted log odds of participants correctly making a negative decision when the correct answer was negative was greater in the delay condition than the deadline condition. The 95% confidence interval does not include zero, so this effect is significant. This indicates that participants were able to avoid false alarms (by making correct rejections) more often in the delay condition than the deadline condition. Figure 3 provides a graphical representation of the interactions between condition, correct answer and number of presentations on the predicted log odds of a negative decision. Neither the condition by

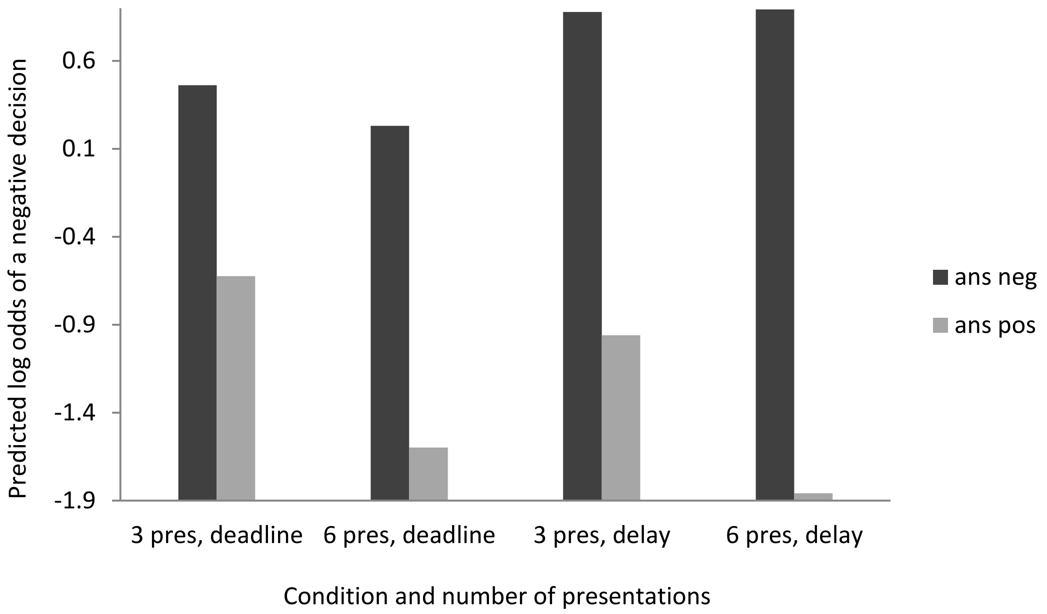

Figure 3. Predicted log odds of a negative decision by correct response, number of presentations, and test condition.

presentations interaction or the three way interaction between condition, presentations and correct answer were significant. For the condition by presentations interaction, this indicates that when the number of presentations changed from three to six, this did not significantly change the effect of condition on the predicted log odds of a negative decision. That is, studying a word more times did not result in less negative decisions in the delay condition, and therefore there was not an increase in positive decisions; this demonstrates that increased presentations did not produce significantly more hits in the delay condition. The three way interaction indicates whether the interaction between correct answer and condition varied depending on number of presentations. Therefore the non-significant regression coefficient indicates that the relationship between correct answer and condition did not vary significantly when presentations increased. Hence participants were more likely to make a correct rejection in the delay condition regardless of number of presentations.

Overall this result demonstrates that participants were able to distinguish between nouns and their plural forms better in the delay condition, as intended, suggesting that recollection was used to a greater degree in this condition. The following analyses investigated whether the confidence-accuracy relationship changed as a result of these changes in recollection.

Confidence and accuracy. Since the key hypotheses for these experiments relate to the effect of the manipulation on negative decisions specifically, the first model investigated the relationship between confidence and accuracy by condition for negative decisions only. Table 10 contains the fixed effects coefficients for this model. The deadline condition was coded as 0 and the delay condition as 1 such that the regression coefficient represents the change between the deadline and delay conditions. This model tested the prediction that there would be an interaction between confidence and condition such that confidence predicted accuracy more strongly in the delay condition. If confidence predicted accuracy more strongly in the

Table 10

Fixed effects coefficients for the mixed effects model predicting accuracy from confidence and condition

Fixed Effect	*b*	SE_b	95% CI
Intercept	-1.94	0.22	[-2.29, -1.54]
Condition (delay)	-0.45	0.32	[-1.06, 0.13]
Confidence	0.04	0.003	[0.03, 0.04]
Condition × Confidence	0.01	0.004	[0.004, 0.02]

Note. The models included test word (*SD* = 0.24) as a random intercept.

delay condition I expected to see a significant positive interaction between confidence and condition.

The interaction between condition and confidence significantly improved the fit of the model as shown by the 95% confidence interval [0.003, 0.02] which does not include zero. The positive value of *b* for the interaction indicates that the positive relationship between confidence and accuracy was stronger in the delay condition, with a weaker confidence-accuracy relationship in the deadline condition, as expected. This interaction is depicted in Figure 4.

To determine whether a positive-negative difference occurred, the relationship between confidence and accuracy was examined for both decision types. A mixed effects model was created with accuracy as the outcome and response, confidence and condition as fixed effects. Table 11 shows the fixed effects coefficients for this model. As the deadline condition was

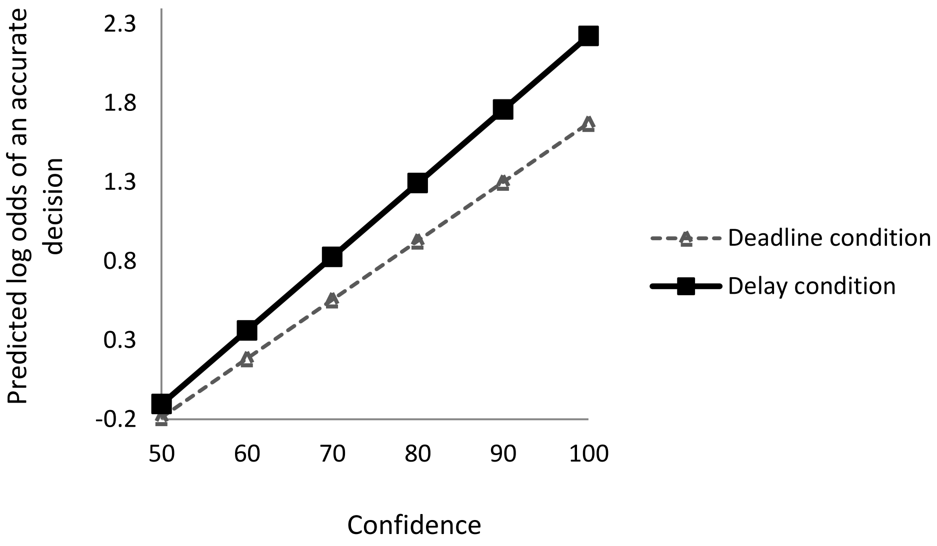

Figure 4. Predicted log odds of an accurate decision by confidence level for negative decisions in each response time condition.

Table 11

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy Overall from Condition, Confidence and Response Type

Fixed Effect	*b*	SE_b	95% CI
Intercept	-2.36	0.22	[-2.57, -2.20]
Condition (delay)	-0.50	0.33	[-0.88, -0.10]
Confidence (conf)	0.03	0.003	[0.03, 0.04]
Response (negative, resp)	0.43	0.31	[-0.06, 0.92]
Condition × conf	0.01	0.004	[0.004, 0.02]
Condition × Resp	0.04	0.46	[-0.63, 0.52]
Conf × Resp	0.001	0.004	[-0.01, 0.01]
Condition × Conf × Resp	0.001	0.006	[-0.01, 0.01]

Note. The models included test word (*SD* = 0.14) as a random intercept.

coded as 0 and the delay condition as 1, the *b* value for the fixed effect of condition indicates the change in accuracy between the deadline condition and the delay condition when confidence was 0 and response was positive (0). Therefore the negative *b* value for condition indicates that for positive decisions when confidence was 0 the delay condition was associated with lower accuracy than the deadline condition. This is not informative as the confidence values in this dataset reflected the confidence ratings participants gave and therefore started at 50. In addition as the 95% confidence interval for this effect includes zero, the effect was not significant.

As in the previous analysis, the *b* value for confidence represents the change in accuracy for one 10% increment increase in confidence, when the other fixed effects (condition and response) are 0. Hence the *b* value of 0.03 for the fixed effect of confidence means that in the deadline condition (coded as 0) for positive responses (also coded as 0) confidence was positively related to accuracy. As the 95% confidence interval for this fixed effect does not include zero, this effect was significant.

As positive decisions were coded as 0 and negative decisions as 1, the regression coefficient for response represents the change in accuracy between positive and negative decisions when the other fixed effects were 0. The *b* value of 0.43 for the fixed effect of response therefore means that in the deadline condition (coded as 0) when confidence was 0, negative decisions were more likely to be accurate. However as the 95% confidence interval for the effect of response includes zero, this effect was not significant.

The condition by confidence interaction indicates whether the confidence-accuracy relationship varied depending on condition. The positive *b* value indicates that when condition changed from deadline (0) to delay (1) the positive confidence-accuracy relationship increased. As the 95% confidence interval for this effect does not include zero, this effect was significant. This indicates that the confidence-accuracy relationship was stronger in the delay condition than in the deadline condition, as would be expected if the deadline condition impaired the confidence-accuracy relationship.

The interaction between confidence and response then indicates the change in the confidence-accuracy relationship when decision type changes from 0 (positive) to 1 (negative). Therefore if the *b* value for this interaction was negative, it would show that the confidence-accuracy relationship was weaker for negative decisions than positive decisions. If the negative interaction between confidence and response was significant, this would show a positive-negative difference. As the regression coefficient for the interaction between confidence and response was positive, this was not the case, and as the confidence interval for the confidence by response interaction includes zero, this interaction was not significant. Therefore a positive-negative difference was not observed.

The three way interaction between response, confidence and condition shows whether the positive-negative difference changed depending on condition. The confidence-accuracy relationship for positive and negative decisions in each condition is shown in Figure 3. As the deadline condition was coded as 0 and the delay condition as 1, the *b* value for this interaction indicates the change in the positive-negative difference from the deadline condition (0) to the delay condition (1). As there was not a positive-negative difference present, I did not expect this interaction to be significant. Consistent with this, the confidence interval for the interaction between confidence, response and condition included zero, showing that there was no change in the positive-negative difference between conditions. As the interaction between condition and confidence was significant and positive, it appears that the confidence-accuracy relationship was stronger in the delay condition for positive, as well as negative decisions. Hence having more recollection available appears to have resulted in a strengthened confidence-accuracy relationship for both decision types. However, as there was no positive-negative difference present, increased use of recollection did not result in changes in the size of the positive-negative difference.

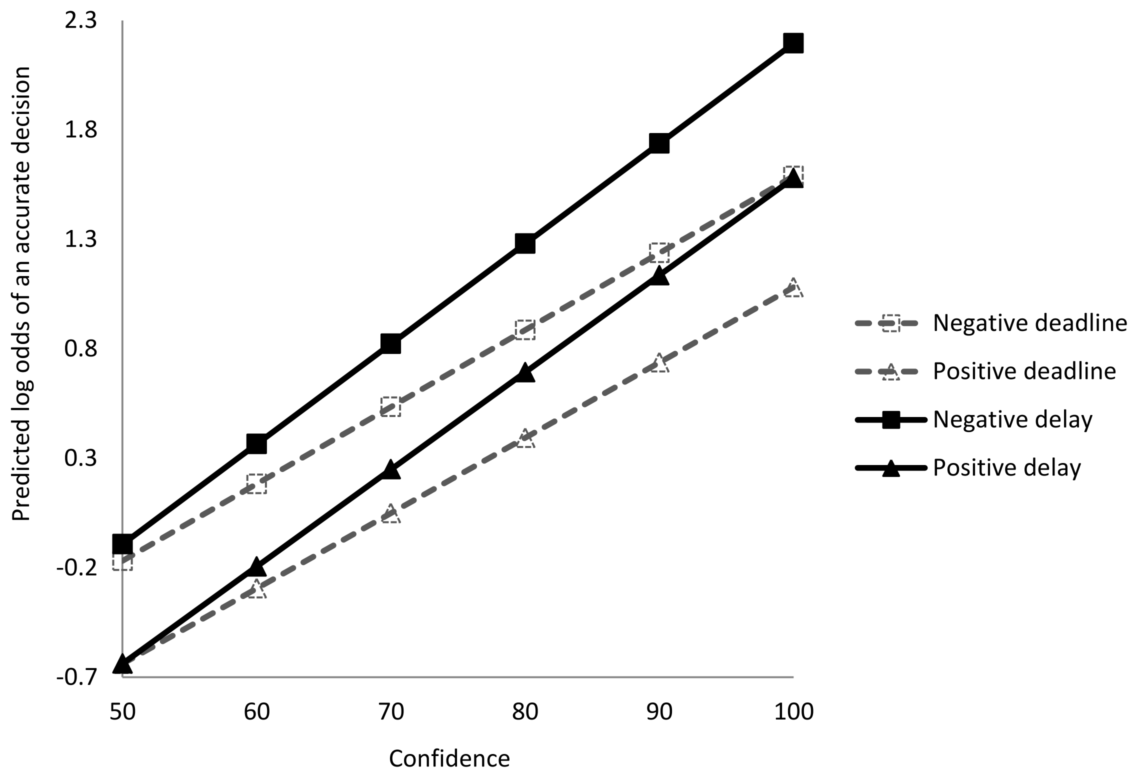

Figure 5. Predicted log odds of an accurate decision by confidence for condition and response type.

Response latency and accuracy. Due to the nature of the manipulation, which affected response latency, conclusions regarding response latency drawn from this experiment are limited in that they assume response latency relationships would follow similar patterns if time allowed for responding was unrestricted, which it is not in this experiment. To investigate this relationship fully a manipulation that does not impact on response latency would need to be used. However the results of analyses on response latency are presented here with figures using the response latency from the time participants were able to respond, to allow for comparison with other experiments where possible. If the response latency-accuracy relationship did change depending on condition this could provide some evidence that recollection may affect the relationship which could be investigated further in later experiments using different manipulations. However these results must be treated with caution as the manipulation altered response latency and therefore may have had an impact on the relationship between response latency and accuracy.

The effect of response latency on accuracy was also first examined for negative decisions only, to determine whether the response latency-accuracy relationship followed the same patterns as confidence-accuracy, as expected. Response latency and condition were added as fixed effects with accuracy as the outcome. The fixed effects coefficients for this analysis are shown in Table 12. The regression coefficient for condition again shows the effect of

Table 12

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy of Negative Decisions from Response-latency and Condition

Fixed Effect	*b*	SE_b	95% CI
Intercept	0.86	0.22	[0.44, 1.38]
Condition, delay (C)	0.59	0.24	[0.02, 0.94]
Response latency (RL)	-0.26	1.66	[-4.42, 2.88]
C×RL	-0.40	1.69	[-3.26, 3.70]

Note. The models included test word (*SD* = 0.29) as a random effect.

condition on the accuracy of negative decisions when response latency is 0; since response latency could not be expected to be 0 this is not informative. The regression coefficient for response latency shows the relationship between response latency and accuracy for negative decisions made in the deadline (0) condition. Although the *b* value is negative, reflecting a negative relationship with accuracy as is the norm for response latency, the 95% confidence interval includes zero, so this effect was not significant.

Because the regression coefficient for response latency is negative, the response latency by condition interaction would need to show a negative relationship to reflect a stronger relationship in the delay condition. That is, a negative *b* value for the interaction would show that the relationship was a stronger negative relationship in the delay condition.

Hence if the response latency-accuracy relationship was stronger in the delay condition, this would be shown by a significant negative interaction between response latency and condition. The interaction between response latency and condition was negative, as expected, however, as the confidence interval for the interaction includes zero, this effect was not significant. Given that the response latency accuracy relationship was not significant in the deadline condition, as shown by the 95% confidence interval for the fixed effect of response latency, this result indicates that there was no significant difference between this and the delay condition; that is, the response latency-accuracy relationship was not significant in the delay condition either. Hence response latency did not predict accuracy in this Experiment. This makes sense as the manipulation specifically affected response latency, and therefore the response latencies recorded do not reflect how long participants took to come to their decision but in a large part reflect how long participants were given to make their decision.

To determine whether there was a positive-negative difference for response latency, the relationship between response latency and accuracy was then examined for both decision types. A mixed effects model was created with accuracy as the outcome and response,

response latency and condition as fixed effects. The fixed effects coefficients for this model are shown in Table 13.

The regression coefficient for the fixed effect of condition shows the change in accuracy for positive decisions (0) when response latency is 0, from the deadline condition to the delay

Table 13.

Fixed Effects Coefficients for the Model Predicting Accuracy from Response Latency, Condition and Response

Fixed Effect	*b*	SE_b	95% CI
Intercept	1.04	0.15	[0.83, 1.19]
Condition (delay, cond)	0.42	0.19	[-0.04, 0.68]
Response latency (L)	-3.99	1.29	[-6.56, -2.87]
Response (negative, resp)	-0.20	0.26	[-0.45, 0.07]
Cond × L	2.35	1.30	[0.71, 5.49]
Cond × Resp	0.17	0.30	[-0.11, 0.43]
L × Resp	3.80	2.05	[2.07, 5.60]
Cond × L × resp	-2.80	2.10	[-4.30, -1.29]

Note. The models included test word (*SD* = 0.28) as a random slope and intercept.

condition. Again, this is uninformative as response latency was never zero. The regression coefficient for the fixed effect of response latency shows the change in the predicted log odds of an accurate decision in the deadline condition (0), for positive decisions (0) when response latency increased by one second. As the *b* value is negative this indicates that the predicted log odds of an accurate decision reduced as response latency increased. The 95% confidence interval does not include zero, so therefore response latency significantly predicted accuracy for positive decisions in the deadline condition.

The *b* value for the effect of response indicates the change in the predicted log odds of an accurate decision in the deadline condition (0) when response latency is 0, between positive and negative decisions. Again this is not useful as response latency cannot be zero. For the response latency by condition interaction, the *b* value indicates the change in the response latency-accuracy relationship between the deadline and delay conditions.

As the response latency-accuracy relationship is negative, the positive value of the regression coefficient for this interaction indicates that the relationship was less strong in the delay condition. As the 95% confidence interval does not include zero, this interaction was significant. Therefore response latency predicted accuracy more strongly in the deadline condition, specifically for positive decisions, as the regression coefficient for response latency refers to positive decisions.

A significant positive response latency by response interaction indicated that a positive-negative difference occurred, with accuracy reducing less as response latency increased for negative decisions. The three way interaction between condition, response and response latency was also significant. This interaction was negative, indicating that there was less difference in the response latency-accuracy relationship between positive and negative decisions in the delay condition. However as Figure 6 depicts, this difference was due to a weaker relationship between response latency and accuracy for positive decisions in the deadline condition, rather than a strengthened relationship for negative decisions. This makes sense as the previous analysis established that response latency did not predict accuracy for negative decisions in either condition. Therefore any change in the positive-negative difference must to be due to a change in the response latency-accuracy relationship for positive decisions.This result suggests that the response latency-accuracy relationship for

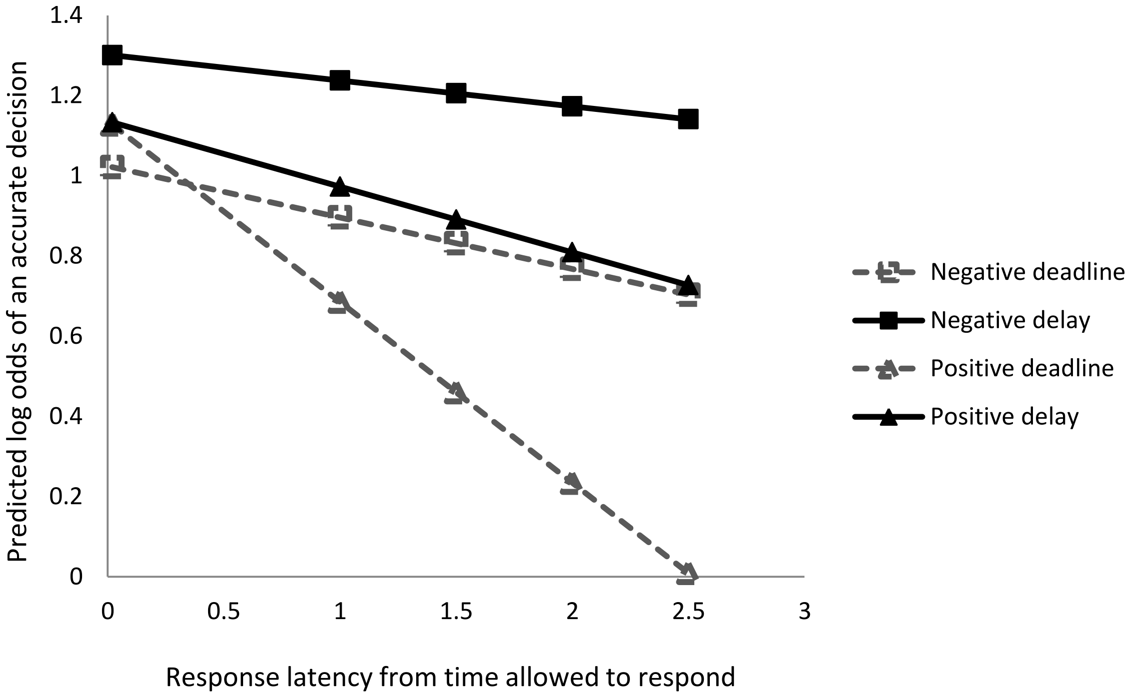

Figure 6. Predicted log odds of an accurate decision by response latency for positive and negative decisions in each condition.

positive decisions may actually be improved by use of a response deadline. However a clear positive-negative difference was demonstrated by the significant interaction between response latency and response. Forcing participants to respond more or less quickly did not affect the response latency-accuracy relationship for negative decisions, for which response latency did not predict accuracy in either condition. On the other hand, for positive decisions, response latency was a stronger predictor of accuracy when participants were forced to respond rapidly, than when they were forced to wait before responding. Hence the results of this experiment do not support the hypothesis that ability to use recollected evidence would strengthen the response-latency accuracy relationship for negative decisions.

General Discussion

These results demonstrate that making more recollected evidence available through enforcing a delay before responding in plurality discrimination decisions can result in a strengthened relationship between confidence and accuracy for negative decisions in comparison with impairing recollection via a response deadline. If the impairment caused by a short response deadline is mainly in recollection, as suggested in the introduction and by previous studies in the literature (e.g., Light et al., 2004) then this provides evidence that the confidence-accuracy relationship is linked to the amount of recollection used in recognition decisions in this task.

The changes in false alarms and hits in the deadline condition compared with the delay condition in Experiment 4 were consistent with those shown in studies where recollection has been impaired (e.g., Light et al., 2004), suggesting that the manipulation was successful in impairing recollection through the short response deadline. Specifically, if both recollection and familiarity were impaired in the deadline condition, a mirror effect would emerge with both false alarms increasing and hits decreasing as demonstrated by a number of investigations of the mirror effect which have been explained using dual-process theory (e.g., Joordens & Hockley, 2000; Malmberg, Holden, & Shiffrin, 2004; Reder, Angstadt, Cary, Erikson, & Ayers, 2000). Because the deadline condition only resulted in increased false alarm rates, without a difference in the hit rate, the data from this experiment suggest that requiring participants to respond rapidly did not impair their ability to recognise recently studied stimuli based on familiarity, but did cause them to falsely recognise similar items due to reduced ability to use recollection.

One alternative explanation for the results that must be considered is that the delay also allows for accumulation of more familiarity evidence, which allows individuals to utilise more evidence when making their decision. However this is unlikely for a number of reasons. First, an increase in familiarity in the delay condition should show up in an increase in hits, which was not evident in this experiment. Second, in single process theories where familiarity only is used, a long latency negative decision, although it may involve more evidence, would still be more difficult due to the amount of evidence for a positive decision lying closer to the criterion for a positive response (e.g., Norman & Wickelgren, 1969). This is likely to decrease confidence, due to the decision becoming more difficult, rather than increase confidence in response to the presence of more familiarity. Hence confidence and accuracy of negative decisions would be expected to be less strongly rather than more strongly related for decisions with longer response latencies. Thus the stronger confidence-accuracy relationship in the delay condition would not be predicted by this explanation. In many single-process theories (e.g., Diller, Nobel, & Shiffrin, 2001; Hintzman, 1988) it is assumed that familiarity is generated relatively immediately as all the memory traces associated with a particular stimulus are activated in parallel. This relates to the third point, which is that most prior research indicates that familiarity should become available and reach its maximum level very rapidly (Basile & Hampton, 2013; Yonelinas & Jacoby, 1994). If this is the case, then all available familiarity evidence should already have been available by the response deadline in the deadline condition, and the delay should not have allowed any further accumulation of familiarity evidence. Hence it is unlikely that the effects of the delay condition resulted from changes in familiarity rather than recollection.

In this plurality discrimination task, there was no significant positive-negative difference evident in the confidence-accuracy relationship. In contrast, in Experiments 2a and 2b, a positive-negative difference did occur. This is supported by the fact that if analysed without the fixed effect of condition, the regression coefficient for the change in the confidence-accuracy relationship in Experiments 2a and 2b combined between positive and negative decisions is -0.19 whereas in this experiment it was a non-significant 0.001[7]. If this is due to a difference in recollection, it would suggest that participants may have utilised recollection more in this experiment. The suggestion that participants used more recollection in Experiment 4 than in Experiments 2a and b combined is supported by the fact that the predicted log odds of a false alarm occurring following increased presentations if analysed without condition as a fixed effect in this experiment was a non-significant 0.06 compared with 0.34 overall in Experiments 2a and 2b. Two differences between the Experiments are possible causes of this discrepancy. First, words were presented more times in Experiment 4, potentially allowing participants to either have higher familiarity associated with all words (e.g., Hintzman, 1988), or have both higher familiarity and greater ease of recollection associated with all words (e.g., Jacoby et al., 1998). A higher level of familiarity for all words could have led to increased presentations increasing familiarity less, or participants may have been more aware of the fact that heightened familiarity was likely to be due to multiple presentations of a similar item. This could have led participants to use recollection more, knowing that all words were likely to be highly familiar, in order to attain a level of accuracy they felt was acceptable in the task (Malmberg, 2008). The multiple presentations may also have made recollection of all words easier due to multiple opportunities for encoding, which may have made recollected evidence more readily available. Second, in Experiment 4, test

[7] Tables of fixed effects for these analyses can be found in Appendix 4A.

words were either studied or changed plurality, with no new words, whereas, in Experiments 2a and 2b, new words were included in the test phase. For new words compared with studied words, familiarity would be an efficient basis for responding as words recently studied could be expected to be more familiar. However for studied words compared with reversed plurality words, familiarity would be less informative as most features of all test stimuli would match information in memory, leading to high familiarity for all test items. The instructions at test in Experiment 4 made it clear that all words would be either as studied or reversed plurality, and thus participants would have been aware that all test stimuli would be highly similar to a studied item. Therefore participants may have realised that familiarity would not be informative. In contrast, in Experiments 2a and 2b, participants would have been aware that some items could be discriminated based on familiarity, and therefore may have used less recollection due to understanding that they could achieve accuracy on at least some items more rapidly without use of recollected evidence. If individuals attempt to complete the task efficiently, and vary the extent to which they use recollection accordingly, as prior research suggests (Malmberg, 2008; Malmberg & Xu, 2007), then the presence of new words in Experiments 2a and 2b may have encouraged less use of recollected evidence than knowledge that all test stimuli would be difficult to discriminate in Experiment 4.

There was no significant response latency-accuracy relationship for negative decisions in this study, and this did not change vary depending on the deadline manipulation. This is likely to be due to the response deadline manipulation controlling response latency to some extent. In addition the results of Experiment 4 could be associated with implementation of a response deadline generally, rather than with changes in recollection. Therefore one aim of subsequent experiments was to find and use a manipulation that did not affect response latency in order to more accurately observe the response-latency accuracy relationship.

As the positive-negative difference did not occur overall in the plurality discrimination task the results of Experiment 4 do not allow direct conclusions about the possibility of reducing or increasing the positive-negative difference through manipulations that affect recollection, as they cannot demonstrate a change in an existing positive-negative difference. The manipulation used in this experiment either impaired or encouraged recollection, but did not compare these conditions to individuals' baseline performance where recollection is neither restricted nor encouraged. Experiment 6 in Chapter 7 aimed to do this, in order to compare a condition in which recollection was used naturally by participants and a condition in which recollection was manipulated. In addition the experiments thus far had focussed on a plurality discrimination task, and hence it was important to test whether the results could be replicated in other types of tasks. Experiment 5 in Chapter 5 served this purpose.

References

Basile, B. M., & Hampton, R. R. (2013). Recognition errors suggest fast familiarity and slow recollection in rhesus monkeys. Learning & Memory, 20(8), 431-437. doi: 10.1101/lm.029223.112

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. Journal of experimental psychology. Learning, memory, and cognition, 27(2), 414-435.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological review, 95(4), 528-551.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. Memory & Cognition, 26(3), 449-462. doi: 10.3758/bf03201155

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. Journal of Memory and Language, 33(1), 1-18. doi: 10.1006/jmla.1994.1001

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. Psychonomic Bulletin & Review, 5(4), 705-709. doi: 10.3758/bf03208849

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. Journal of experimental psychology. Learning, memory, and cognition, 26(6), 1534-1555.

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. Memory & Cognition, 32(7), 1182-1193. doi: 10.3758/bf03196891

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. Cognitive Psychology, 57, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. Journal of Experimental Psychology-Learning Memory and Cognition, 30(2), 319-331. doi: 10.1037/0278-7393.30.2.319

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. Memory & Cognition, 35(3), 545-556. doi: 10.3758/bf03193293

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. Journal of Mathematical Psychology, 6(2), 192-208. doi: 10.1016/0022-2496(69)90002-9

Reder, L.M., Angstadt, P., Cary, M., Erikson, M.A. , & Ayers, M.A. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. Learning, memory, and cognition, 26, 294-320. doi: 10.1037/0278-7393.26.2.294

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory - effects of interference and of response speed. Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 48(4), 516-535. doi: 10.1037/1196-1961.48.4.516

Appendix 4A: Fixed Effects coefficients for analyses without condition as a fixed effect in Experiment 4

Table A1.

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy From Response and Confidence for Experiment 4

Fixed Effect	*b*	SE_b	95% CI
Intercept	-2.61	0.17	[-3.06, -2.26]
Confidence (Conf)	0.04	0.002	[0.035, 0.044]
Response (negative, resp)	0.45	0.23	[0.28, 0.88]
Conf×Resp	0.001	0.003	[-0.004, 0.003]

Note. The model included test word (SD = 0.13) as a random effect.

Table A2.

Fixed Effects Coefficients for the Mixed Effects Model Predicting False alarms from Number of Presentations in Experiment 4

Fixed Effect	*b*	SE_b	95% CI
Intercept	-0.73	0.09	[-0.88, -0.58]
Presentations	0.06	0.08	[-0.07, 0.24]

Note. The models included test word (*SD* = 0.27) and participant (*SD* = 0.61) as random intercepts and random slopes for presentations.

CHAPTER 5 — RESTRICTING RECOLLECTION: AUDIO VERSUS TEXT TASK

To this point this thesis has used two methods to test the idea that neglect of recollected evidence causes the positive-negative difference. First, I compared an item recognition task, in which very little recollected evidence is available to participants, with an associative task (plurality discrimination) in which recollected evidence is critical to accuracy. Next, the plurality discrimination task was used again, this time implementing a response time manipulation to restrict recollection in the deadline condition and encourage it in the delay condition. These experiments found some evidence that the positive-negative difference occurs in word recognition and plurality discrimination and that the confidence-accuracy relationship for negative decisions may be impaired when a recollected evidence is restricted, in this case when a task provides less recollected evidence or when a deadline is implemented for responding at test. However, since the only associative task used in the first set of experiments was plurality discrimination, the extent to which manipulation of recollection would be effective in changing the confidence-accuracy relationship in other tasks is uncertain. In Experiment 5 I aimed to test the generalisability of this effect by using the response deadline manipulation in a different associative task. In addition, this provided an opportunity to observe whether the positive-negative difference occurs in another type of associative task, as would be expected if the cause is underlying levels of recollected evidence.

Experiment 5

For this experiment I used a task which was developed by Jacoby, Jones, and Dolan (1998) to test for the presence of recollected evidence in participants' decisions by placing familiarity and recollection in opposition. Jacoby et. al. used this task to demonstrate use of recollection when a delay is enforced and reliance on familiarity when response time is restricted. They were able to demonstrate that false alarms increased with greater memory strength (more repetitions) of an item, when response time was restricted, but decreased with greater memory strength when a delay before response was enforced. This supports the notion that familiarity is relied upon in decisions that are made very rapidly. In Experiment 5 I attempted to replicate this procedure while introducing a confidence judgment, to observe the effect of restricting and encouraging recollection on the confidence-accuracy relationship. As in Experiment 4, based on Jacoby et al.'s work, the deadline condition was expected to restrict the amount of recollected evidence available, while the delay condition would enforce time for participants to utilise recollected evidence. If failure to optimally utilize recollection is responsible for the positive-negative difference, negative decisions should display poorer confidence-accuracy relationships in the deadline condition, and strengthened confidence-accuracy relationships in the delay condition.

The overall confidence-accuracy relationship and the extent to which a positive-negative difference occurred were also of interest. Although the main test of the proposed theory is the impact of the deadline on negative decisions specifically, for this effect to have ideal practical applications it would also need to show an effect of increasing the difference between positive and negative decisions, such that the positive-negative difference can be altered overall. This would lead to predictions that if recollection were enhanced, rather than impaired, the accuracy of recognition decisions could be estimated based on confidence, regardless of the type of decision made. It is possible that delaying responding would result in increased recollection being used, and hence an additional question was whether the enforced delay could be used to enhance the confidence-accuracy relationship. However, some dual-process theories suggest that individuals will not attempt to use recollection unless familiarity is high enough to suggest it will help them decide if an item is studied (e.g., Malmberg, 2008). Therefore it was also possible that participants would not use recollection in low-familiarity decisions even if the delay made it available. Hence the main predictions for this experiment involved the negative effect of the deadline condition on recollection.

I predicted that the deadline condition would result in a weaker confidence-accuracy relationship for negative decisions than the delay condition. As explained in Chapter 4, much prior research suggests that recollection takes longer to become available than familiarity (e.g., Hintzman & Curran, 1994; McElree, Dolan, & Jacoby, 1999; Yonelinas & Jacoby, 1994) and that therefore less recollected evidence is available when a short response deadline is applied. In addition Experiment 4 provided evidence that this occurs in plurality discrimination, and thus I expected that it would also occur in other tasks where all items at test are similarly familiar. I also predicted that a positive-negative difference would be present in the deadline condition, due to lack of available recollected evidence to support confidence ratings in negative decisions. As it was not certain whether the delay condition would increase participants' use of recollection, I did not predict whether the confidence-accuracy relationship would show an interaction between condition and response type. The delay condition could be expected to increase the availability of recollection, but not necessarily participants' use of this evidence for negative decisions. If the delay condition did increase participants' use of recollected evidence, I expected the positive-negative difference to be reduced in the delay condition compared with the deadline condition.

Method

Participants. Participants were 80 Flinders University students (59 female) between the ages of 18 and 60 (*M* = 22.1, *SD* = 5.86). All participants had English as a first language, normal or corrected-to-normal vision, and normal or corrected-to-normal hearing.

Design. The experiment used a 2 (repetitions: one, three) × 3 (stimulus type: heard stimulus, read stimulus, new stimulus) × 2 (test type: delay [must wait for 2 *s*, then respond as quickly as possible], deadline [must respond within 2 *s* after word appears]) mixed design. Repetitions and stimulus type were varied within-subjects, and test type was varied between subjects. The dependent variables were confidence judgments (50-100%) about each decision, response latency, and accuracy of each decision.

Materials. Stimuli were 192 words, with equivalent ranges for word length, word frequency, and concreteness based on the same ranges as for Experiment 4, obtained from the University of Western Australia's MRC Psycholinguistic Database (Wilson, 1988). Words were randomly allocated to be presented in read or heard form (a recorded audio version and a text version of every word was available, and the experimental software randomized which was presented) and to be presented once or three times, for each participant. The read and heard lists were presented interspersed. At study words were displayed in the centre of a computer screen for text words, and played through the headphones for audio words. At test all words were presented in text form in the centre of the computer screen. All audio words were spoken by a young female, recorded on digital media and converted to WAV format for inclusion in the experimental software.

Procedure. Participants read a short introduction to the experiment explaining what was involved in the task before giving informed consent. Participants completed the experiment on computers in individual rooms. Instructions and written stimuli appeared on the computer screen, with each section on a new screen which participants reached by pressing the space bar on the keyboard. The experiment was in six study-test blocks. In the study phase of each block, participants were instructed that they would study a series of words, some of which would be played through the headphones and others of which would appear on the screen. They were told that they should try to remember the words as they would later be asked questions about them, and notified that some words would appear multiple times. When they indicated that they were ready to begin, participants studied 24 words one at a time, half of which were played through the headphones, and half of which appeared on the screen. Text words were displayed for 2 *s* each and a 2 *s* inter-stimulus interval was used. Words were randomly allocated for each participant to blocks, presentation modalities and number of presentations. In each block, half of the words were presented once and the other half three times.

In the test phase, a prompt displayed on the screen above the test word, which was presented in the centre of the screen, asking either *was this text?* or *was this voice?* Participants' task was to indicate whether the word on screen had been studied in the modality suggested. Hence, if the prompt said *was this voice?* they were only to respond *yes* if they believed they had heard that word through the headphones. If the word had been presented on screen, or had not been studied, the correct response was *no*. Participants used the keyboard to indicate their decision, pressing the *Z* key for *yes* and the *?* key for *no*. These keys were marked with stickers that had *yes* and *no* printed on them as appropriate. Following their recognition decision, participants were asked to indicate their confidence that their decision was correct, by pressing a key on the keyboard from 5 (50%) to 0 (100%). These keys were marked with stickers printed with the appropriate confidence values.

Participants were randomly allocated to either the deadline test or delay test condition. In the deadline condition, participants were told that they must respond as quickly as possible after the word and prompt appeared on screen. If they took more than 2 *s* to respond, the word disappeared and they were required to move on to the next screen and provide a confidence rating, which they were told should be 50% if they had not given an answer. In the delay test condition, participants were told that after the word and prompt appeared on the screen, they would be required to wait until the *yes* and *no* buttons appeared on screen before giving their answer. Once the buttons appeared, they were required to respond as quickly as possible. For the first 2 *s* of each test trial, the test word and prompt were displayed without the response buttons. After 2 *s* had passed, the response buttons appeared and participants could select a response. Once they had completed all 24 trials for the first block, participants moved on to a new study list. This procedure was continued until participants had finished all six blocks.

Results

Manipulation checks. To test whether the deadline manipulation produced changes in recollection, a mixed effects model was created with condition and number of presentations as fixed effects and false alarms as the outcome. The fixed effects coefficients for this model are shown in Table 13. If the deadline condition impaired use of recollection, false alarms should increase with number of presentations in the deadline condition, while increasing less or remaining stable with presentations in the delay condition. This was expected to be shown by a significant interaction between condition and presentations, such that false alarms increased significantly more with increasing presentations in the deadline condition than in the delay condition. Hits were expected to increase with both increasing recollection and increasing familiarity, and therefore to increase with number of presentations in both conditions.

The model predicting false alarms demonstrated a significant interaction between test condition and number of presentations, as expected, as shown by the 95% confidence interval for the interaction which does not include zero. The negative value of *b* for the interaction indicates that the effect of presentations on false alarms was significantly lower in the delay condition (coded as 1) than the deadline condition (coded as 0). This interaction is depicted in Figure 7. This result suggests that increasing presentations of a similar stimulus made it easier for participants to distinguish from its modality-altered counterpart in the delay condition, but in the deadline condition increasing presentations prompted more positive responses to similar stimuli. This suggests a higher influence of recollected evidence in the delay condition as intended.

Table 14

Fixed Effect Coefficients for Logistic Mixed-Effects Model Predicting False Alarms and Hits from Number of Presentations and Test Condition

Fixed effect	*b*	SE_b	*95% CI*
	False alarms		
Intercept	-0.81	0.14	[-1.23, -0.38]
Presentations (3, P)	0.12	0.11	[-0.10, 0.33]
Condition, (delay, C)	- 0.76	0.20	[-1.13, -0.38]
P×C	-0.47	0.17	[-0.79, -0.15]
	Hits		
Intercept	0.34	0.10	[0.23, 0.45]
P	1.13	0.08	[0.97, 1.30]
C	0.42	0.14	[0.15, 0.69]
P×C	0.17	0.12	[-0.08, 0.41]

Note: Model included test word (*SD* = 0.27) as a random slope and intercept.

In a similar model predicting hits (fixed effect coefficients also shown in Table 14), there was no significant interaction between number of presentations and test condition, as shown by the confidence interval which includes zero. This indicates that the effect of number of presentations did not differ significantly depending whether the test was under deadline or delay conditions. A higher rate of false alarms with increasing presentations in the deadline condition with no difference in hits supports the idea that recollection occurred to a lesser

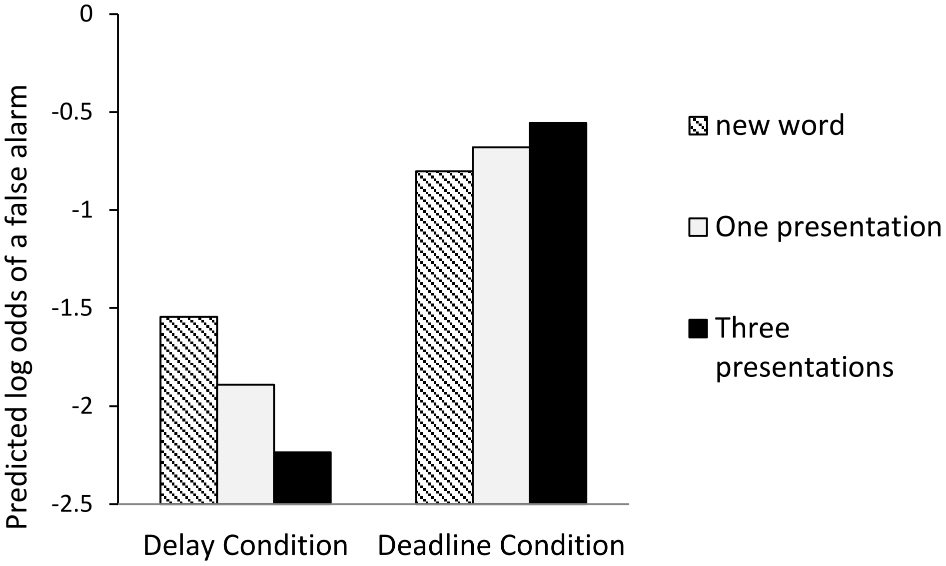

Figure 7. Predicted log odds of a false alarm by test condition and number of presentations.

extent in the deadline condition than in the delay condition. If recollection also increased with presentations, a mirror effect should emerge with hits being lower as well as false alarms higher in the deadline condition than the delay condition (Cary & Reder, 2003; Reder, Angstadt, Cary, Erikson, & Ayers, 2000). The hits portion of this mirror effect is attributed to recollection (Cary & Reder, 2003). Only the false alarm portion occurred in Experiment 5, which is attributed to familiarity. Thus the pattern of false alarms and hits found suggests that that recollection was used in the delay to avoid false alarms, but was not necessarily increased in the delay condition compared with the extent to which it would normally be used. Hence the response deadline impaired recollection but the delay before responding merely allowed recollection rather than increasing recollection.

Table 14 displays the fixed effects coefficients from a model created to test whether response bias differed between the deadline conditions and number of presentations. As positive

Table 14

Fixed Effects Coefficients for the Mixed Effects Model Predicting Response from Correct Answer, Condition and Number of Presentations

Fixed Effect	*b*	SE_b	95% CI
Intercept	-0.30	0.05	[-0.40, -0.20]
Answer (A, negative)	1.02	0.09	[0.82, 1.31]
Condition (cond, delay)	-0.41	0.07	[-0.53, -0.22]
Presentations (6, P)	-1.07	0.08	[-1.24, -0.88]
A × Cond	1.11	0.13	[0.72, 1.27]
A × P	0.97	0.13	[0.56, 1.33]
Cond × P	-1.16	0.12	[-0.42, 0.03]
A × Cond× P	0.58	0.20	[0.13, 1.25]

Note. The models included test word (*SD* = 0.22) as a random slope and intercept.

responses were coded as zero and negative responses as one, the regression coefficients represent the change in the predicted log odds of a negative decision between the levels of each fixed effect if the other fixed effects are zero. Hence the fixed effect of answer indicates the change in the predicted log odds of a negative decision when the correct answer changes from being positive to being negative, for the single presentation and deadline conditions. As the *b* value is positive it indicates that participants were more likely to make a negative decision when the correct answer was negative, in the deadline condition when words had only been studied once. This effect is significant as the 95% confidence interval does not include zero. Thus participants were able to discriminate between target and foil items on the test even in the most difficult conditions.

For the fixed effect of condition, the negative regression coefficient demonstrates that participants were less likely to make a negative decision in the delay condition when the correct answer was positive and the word had been studied once. This effect was also significant. Hence participants were more likely to make a correct negative decision in the delay condition than the deadline condition when words had been presented once.

The fixed effect of presentations indicates the change in the predicted log odds of a negative decision between the single presentation and three presentation conditions for the deadline condition when the correct answer was positive. The negative regression coefficient indicates that an incorrect negative decision was less likely in the deadline condition following three presentations than following a single presentation. This effect was also significant as indicated by the 95% confidence interval.

For the interaction between correct answer and condition, the regression coefficient is positive, indicating that in the delay condition compared with the deadline condition, the effect of correct answer (that a negative decision was more likely when the correct answer was negative) was stronger. This effect was also significant, as shown by the confidence interval, indicating that correct negative decisions were more likely in the delay condition than the deadline condition. The *b* value for the correct answer by presentations interaction is also positive, meaning that the increase in the predicted log odds of a negative decision when the correct answer was negative was larger following three presentations than one presentation. This effect was also significant. The condition by presentations interaction was not significant, which indicates that the change in the log odds of a negative decision between one presentation and three presentations was not significantly larger or smaller in the delay condition compared with the deadline condition. Hence increased presentations led to more accurate negative decisions in both conditions.

The three way interaction between correct answer, presentations and condition was significant. The positive regression coefficient indicates that the positive interaction between correct answer and condition was stronger following three presentations than following one presentation. Thus correct negative decisions were more likely in the delay condition and this effect was stronger when words had been studied three times. This reflects the fact that false alarms occurred more following three presentations in the deadline condition, and thus correct rejections occurred less.

A mixed effects model with response as the outcome and only condition as a fixed effect, for which fixed effects coefficients are shown in Table 15, revealed that there was no significant effect of condition on response. Hence the delay condition showed increased accuracy due to fewer false alarms, but this was not due to a bias to give more negative responses.

Table 15

Fixed Effects Coefficients for the Mixed Effects Model Predicting Response from Condition in Experiment 5

Fixed Effect	*b*	SE_b	95% CI
Intercept	-0.10	0.03	[-0.15, -0.03]
Condition (Delay)	0.03	0.04	[-0.03, 0.09]

Note. The model included test word (*SD* = 0.15) as a random slope and intercept

Confidence. To test the hypothesis that confidence and accuracy would be more strongly related for negative decisions made in the delay condition than those made in the deadline condition, a mixed effects model was created for negative decisions only, with confidence and condition as fixed effects and accuracy as the outcome. The fixed effects coefficients for this analysis are shown in Table 16. If a positive confidence-accuracy relationship existed for

Table 16

Mixed Effects Coefficients for Confidence-Accuracy and Response Latency-Accuracy Relationships for Negative Decisions by Deadline Condition

Fixed Effect	*b*	SE_b	*95% CI*
	Confidence		
Intercept	-1.34	0.16	[-1.58, -1.15]
Condition, (delay, C)	-0.49	0.25	[-0.99, 0.02]
Confidence (conf)	0.02	0.002	[0.02, 0.03]
C × Conf	0.01	0.003	[0.01, 0.02]
	Response latency		
Intercept	0.17	0.20	[0.003, 0.53]
C	1.17	0.22	[0.74, 1.61]
Response latency (RL)	2.38	1.41	[-0.26, 5.09]
C × RL	-3.31	1.42	[-6.08, -0.63]

Note: The models included test word (*SD* = 0.25) as a random slope and intercept.

decisions made in the deadline condition, this would be shown by a significant positive *b* value for the fixed effect of confidence. As the deadline condition was coded as zero and the delay condition as one, a stronger confidence-accuracy relationship in the delay condition would then be reflected by a significant positive regression coefficient for the interaction between confidence and condition. This would indicate that the positive relationship between confidence and accuracy that existed in the deadline condition was stronger in the delay condition.

The fixed effect of condition was not significant as shown by the confidence interval which does not include zero. This indicates that there was no significant difference in accuracy between the deadline and delay conditions when confidence was zero. This is not informative for this dataset as confidence was recorded using the values participants selected, and this 50 was the lowest confidence value. For the fixed effect of confidence, the significant positive regression coefficient shows that in the deadline condition (coded as zero) there was a positive relationship between confidence and accuracy such that increased confidence was associated with higher accuracy, as expected.

Consistent with the hypothesis, there was a positive interaction between confidence and condition, as shown by the positive value of *b*, which was statistically significant as shown by the confidence interval which does not include zero. The *b* value of 0.01 for the interaction indicates that for each 10% increase in confidence the predicted log odds of an accurate decision increased .01 more for decisions made in the delay condition than decisions made in the deadline condition.

A similar analysis was then conducted with response latency[8] as a fixed effect to determine whether the response latency-accuracy relationship followed similar patterns. As the manipulation altered response latency and these data are based on participants' reactions

[8] Response latencies were rounded to four decimal places to satisfy the requirement for consistent value lengths for the mixed effects model.

from the time they were allowed to select a response, they may not represent individuals' natural manner of responding and as such conclusions here are limited. However these data were analysed for comparison with the confidence-accuracy relationship and with previous work. The fixed effects coefficients for this analysis are also presented in Table 16.

If a negative response latency-accuracy relationship existed in the deadline condition this would be shown by a significant negative *b* value for the fixed effect of response latency. If this relationship was stronger in the delay condition, this would then be shown by a significant negative regression coefficient for the interaction between response latency and condition. This would show that the negative relationship between response latency and accuracy that existed in the deadline condition was stronger (more negative) in the delay condition.

For the fixed effect of condition the significant positive *b* value indicates that decisions made in the delay condition were more accurate than those in the deadline condition when confidence was zero; which again is uninformative due to the value of confidence not being within the range in the dataset. The fixed effect of response latency was not significant, indicating that for negative decisions in the deadline condition there was no significant response latency-accuracy relationship.

The test condition by response latency interaction was significant, as demonstrated by the confidence interval for the interaction which does not include zero. The direction of the interaction was negative, as shown by the negative value of *b*. This indicates that the response latency accuracy relationship, from being non-significant in the deadline condition, changed to be a significantly more negative relationship in the delay condition. Figure 8 depicts this interaction, and shows that negative decisions only showed the usual negative relationship in the delay condition, appearing to have a positive response latency-accuracy relationship in the deadline condition.

Overall this result could be taken to support the prediction that response latency would

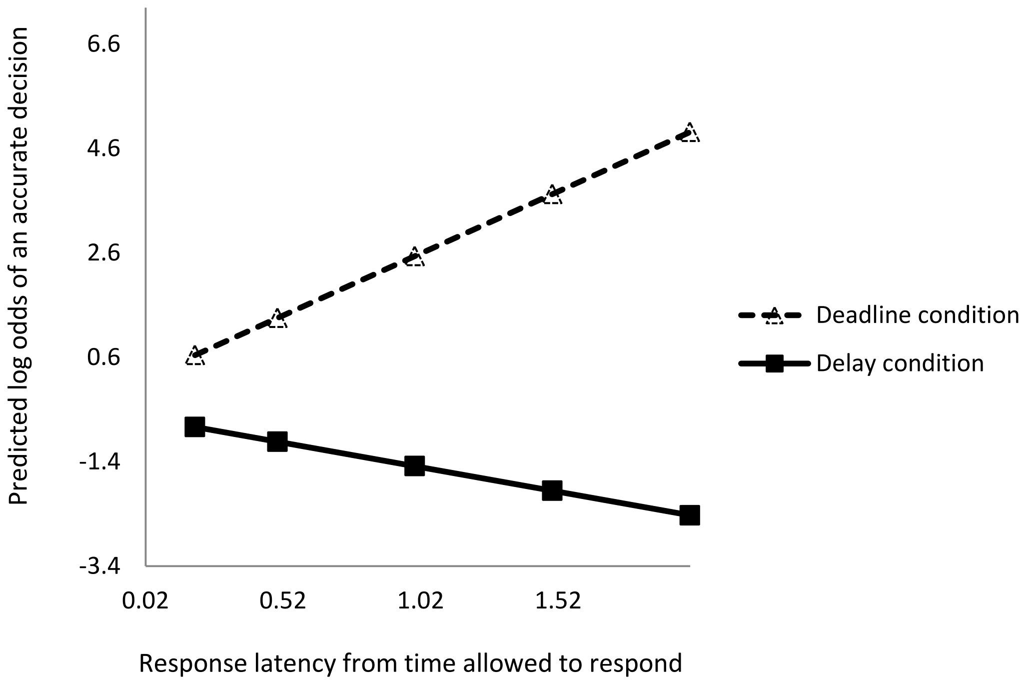

Figure 8. Predicted log odds of an accurate decision by response latency from time allowed to respond and test condition.

follow similar patterns to confidence with a stronger negative response-latency-accuracy relationship in the delay condition, since a negative relationship between response latency and accuracy, similar to that normally seen for positive decisions, could only be seen in the delay condition for negative decisions. However these results must be treated with caution due to the nature of the manipulation.

The positive-negative difference. Both the confidence-accuracy and the response-latency-accuracy relationships were also analysed with decision type included as a fixed effect to indicate whether a positive-negative difference occurred and whether the delay condition reduced this difference. The fixed effects coefficients for both analyses are presented in Table 17.

Table 17

Fixed Effect Coefficients for Logistic Mixed-Effects Model Predicting accuracy overall from Confidence, Condition and Response and Response Latency, Condition and Response

Fixed Effect	*b*	SE_b	*95% CI*
	Confidence		
Intercept	-2.61	0.20	[-3.01, -2.21]
Condition, (delay, C)	-0.30	0.32	[-0.88, 0.03]
Confidence (Conf)	0.05	0.002	[0.04, 0.05]
Response, negative (R)	1.28	0.25	[0.83, 1.52]
C × conf	0.01	0.004	[0.01, 0.02]
C × R	-0.20	0.40	[-0.67, 0.40]
Conf ×R	-0.02	0.003	[-0.02, -0.02]
C × resp × conf	0.002	0.01	[0.01, 0.01]
	Response Latency		
Intercept	1.82	0.21	[1.54, 2.06]
C	0.38	0.23	[0.14, 0.49]
Response latency (RL)	-4.96	1.48	[-6.24, -4.35]
R	-1.67	0.29	[-1.97, -1.39]
C × RL	3.66	1.50	[2.98, 5.12]
C × resp	0.81	0.32	[0.36,1.23]
R × RL	7.47	2.06	[5.29, 9.67]
C × R ×RL	-7.10	2.08	[-9.39, -4.79]

Note. The models included test word (*SD* = 0.22) as a random slope and intercept.

For the analysis of confidence, a positive-negative difference would be shown if a positive confidence-accuracy relationship existed for positive decisions in the deadline condition, and the confidence by response interaction was significant with a negative regression coefficient. This would indicate that the positive relationship observed for positive decisions was weaker for negative decisions. The effect of condition on this relationship would then be shown by the three way interaction between confidence, response and condition. If the delay condition reduced the positive-negative difference, the regression coefficient would be positive, showing that the negative impact of a decision being negative on the confidence-accuracy relationship was smaller in the delay condition.

The confidence by response interaction was significant, as shown by the confidence interval for the response by confidence interaction which does not include zero. The negative *b* value indicates that the confidence-accuracy relationship was weaker for negative decisions, demonstrating a positive-negative difference. The three way interaction between confidence, response and condition did not significantly predict accuracy, as shown by the confidence interval including zero, demonstrating that test condition did not change the size of the positive-negative difference for confidence:

The significant positive interaction between condition and confidence indicates that for positive decisions, the confidence-accuracy relationship was stronger in the delay condition than in the deadline condition overall. Hence the delay condition improved the confidence-accuracy relationship for positive decisions. The previous analysis of negative decisions demonstrated that the delay condition also increased the confidence-accuracy relationship for negative decisions. Thus, the lack of a three way interaction between condition, confidence and response is due to the fact that the confidence-accuracy relationship for both positive and negative decisions benefitted from the delay condition similarly. In Figure 9 this interaction is depicted.

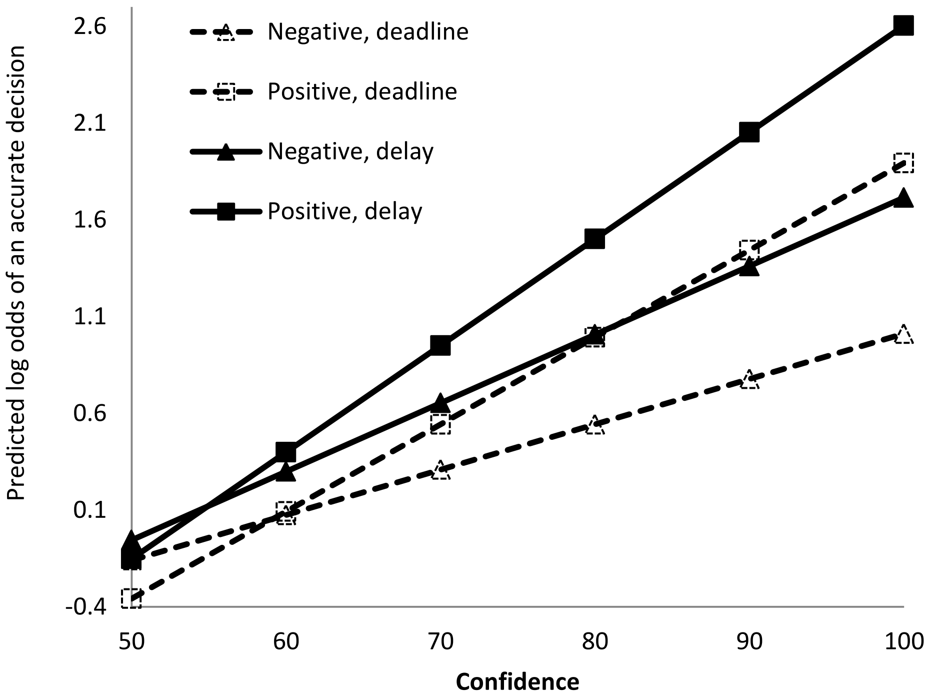

Figure 9. Predicted log odds of an accurate decision by test condition, confidence and response.

For the analysis of response latency the fixed effect of response latency was significant with a negative regression coefficient indicating that for positive decisions in the deadline condition faster responses were more accurate. The response by response latency interaction was also significant, and the *b* value was positive, indicating that in the deadline condition, the negative response latency-accuracy relationship for positive decisions was weaker for negative decisions. The three-way interaction between response latency, response and condition was significant, as demonstrated by the confidence interval which does not include zero. The negative value of *b* indicates that the positive-negative difference was smaller in the delay condition than the deadline condition. However, as Figure 10 makes evident, this was due to the fact that for negative decisions, the response latency-accuracy relationship was positive in the deadline condition. Positive decisions showed a stronger negative relationship

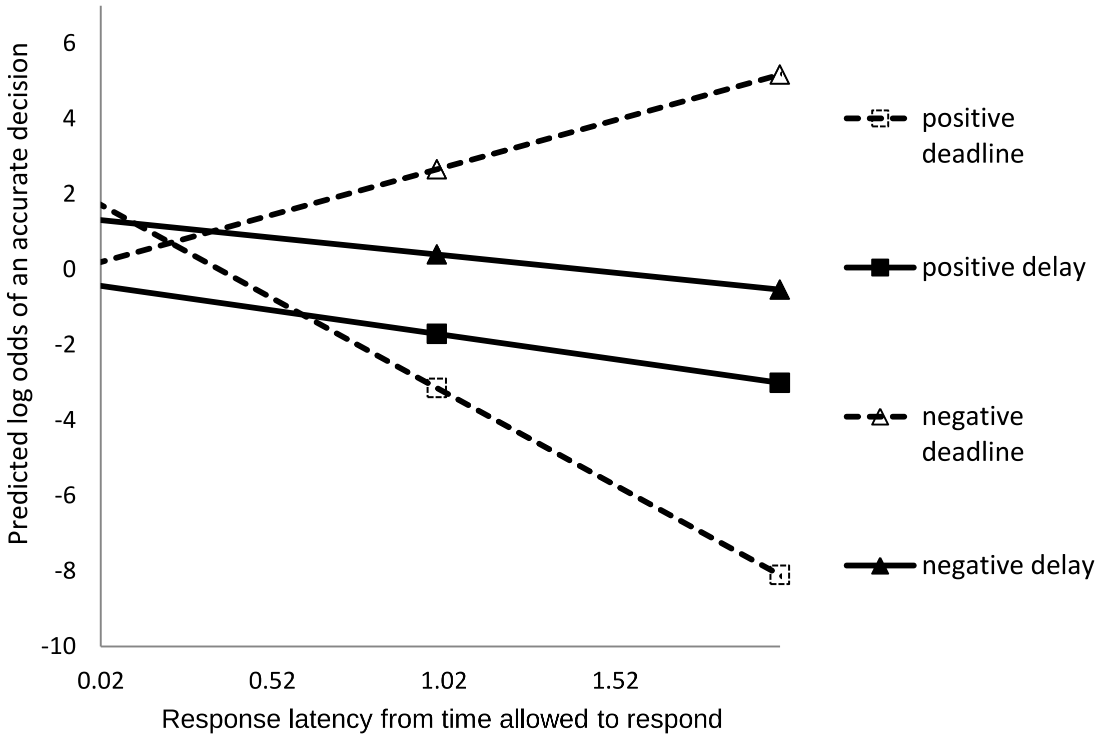

Figure 10. Predicted log odds of an accurate decision by response latency from time allowed to respond and test condition.

than negative decisions in the deadline condition, due to negative decisions showing a positive relationship under deadline conditions. This result suggests quite different processes underlying positive and negative decisions, but again, must be treated with caution for this task. The response latency-response interaction was significant, but was in the direction of negative decisions having a stronger response-latency-accuracy relationship, as shown by the positive value of *b* for the interaction.

Discussion

This experiment tested the idea that the poor relationship between accuracy and confidence for negative recognition decisions results from neglect of recollected evidence.

The response deadline impaired participants' ability to use confidence to indicate their accuracy for both positive and negative decisions. As positive and negative decisions were affected similarly by the response deadline, the results suggest that the confidence-accuracy relationship for positive decisions also benefits from additional use of recollection. This is intuitively likely since providing additional evidence for a decision should allow individuals to better assess how much support each decision has. This assessment can then guide individuals to make more informative confidence judgments based on the amount of evidence for their decision. In addition, it could still be expected that recollected evidence may be easier to retrieve for positive than negative decisions under conditions where it is available, as the target item being viewed at test may act as a stronger cue for retrieval of recollected information than a similar item which has less matching features to cue retrieval. Hence positive and negative decisions may not have been equal in the amount of recollected evidence available even in the delay condition.

Since the main question concerned whether changes in recollection affected the confidence-accuracy relationship for negative decisions, the results did support the idea that recollection is influential in determining the strength of this relationship. As the availability of recollection did not alter the size of the positive-negative difference the results of this experiment did not suggest whether encouragement of recollection might have practical applications for reducing the positive-negative difference. This was the focus of Experiment 6. However, where recollected evidence was available the confidence-accuracy relationship for negative decisions did improve in comparison with when recollected evidence was restricted, and this means that where recollected evidence is used, confidence may be sensibly used to assess the accuracy of negative decisions to some degree, perhaps even in tasks where this is not normally possible. Experiment 5 demonstrated that the effects found in Experiment 4 can generalise to a different recognition task, which suggests that the amount of recollection available for decisions may impact on the confidence-accuracy relationship in a variety of recognition tasks. As both Experiment 4 and Experiment 5 used a response deadline to manipulate the availability of recollected evidence, however, a task for subsequent experiments was to test whether this effect occurred with different manipulations.

One issue which must be considered in relation to the confidence results is that increases in accuracy overall can produce an increase in the confidence-accuracy relationship (e.g., Perfect & Stollery, 1993). The delay condition did result in increased accuracy overall as demonstrated by the reduction in false alarms without reduction in hits. This specific pattern of increased accuracy is consistent with an increase specifically in recollection (e.g., Reder et al., 2000). Jacoby et al. (1998) also interpreted a similar effect produced by the same manipulation as a specific effect on recollection. It is possible that increased use of recollection has its effect on the confidence-accuracy relationship through an increase in overall accuracy that it produces. Accuracy could be expected to increase overall when participants have access to more evidence for their decisions, and hence use of recollected evidence and increasing overall accuracy are difficult to separate. This is only problematic for the influence of recollection on the confidence-accuracy relationship if increased use of recollection can only have its effect when memory quality is high enough for overall accuracy to increase. Thus a task for future studies to test the effect of increased use of recollected evidence in tasks where participants' memory quality does not allow for optimal accuracy. This may be the case in real-world recognition situations such as eyewitness identification where study time may be limited or viewing conditions less than optimal (e.g., Palmer, Brewer, Weber, & Nagesh, 2013). If availability of recollection affects the relationship between confidence and accuracy for negative decisions, even when memory quality is low, then these findings will have more applicability to real world situations where recognition decisions are required.

This experiment also supported the idea that the response latency-accuracy relationship follows similar patterns to the confidence-accuracy relationship with the response latency-accuracy relationship for negative decisions being negative when recollected evidence was more available. In contrast in the deadline condition the response latency-accuracy relationship for negative decisions was non-significant and trended towards a positive relationship. Conclusions concerning response latency in this study are limited by the fact that the manipulation controlled response latency, and hence participant's response times are not representative of natural responding. Therefore the results concerning response latency should be treated with caution.

A logical next step was to design a manipulation which would affect recollection, either through impairment or encouragement, without artificially altering response latency. This would allow clearer conclusions concerning the response-latency-accuracy relationship, and ensure that the effects found in this experiment and in Experiment 4 are not caused in some way by manipulation of response time that is unrelated to recollection. In Chapters 6 and 7 I report experiments in which this was attempted.

References

Cary, M. , & Reder, L.M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. Journal of Memory and Language, 49(2), 231-248.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. Journal of Memory and Language, 33(1), 1-18. doi: 10.1006/jmla.1994.1001

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. Psychonomic Bulletin & Review, 5(4), 705-709. doi: 10.3758/bf03208849

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. Cognitive Psychology, 57, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. Journal of Experimental Psychology-Learning Memory and Cognition, 25(3), 563-582. doi: 10.1037//0278-7393.25.3.563

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. Journal of Experimental Psychology-Applied, 19(1), 55-71. doi: 10.1037/a0031602

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults - one deficit or two? Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46(1), 119-135.

Reder, L.M., Angstadt, P., Cary, M., Erikson, M.A. , & Ayers, M.A. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. Learning, memory, and cognition, 26, 294-320.

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory - effects of interference and of response speed. Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 48(4), 516-535. doi: 10.1037/1196-1961.48.4.516

CHAPTER 6 — EXPERIMENTS WITH DIVIDED ATTENTION

In previous chapters I tested the idea that neglect of recollected information causes the positive-negative difference, first by comparing tasks in which information available to be recollected is minimal versus necessary, and then by attempting to manipulate the amount of recollected evidence available for decisions. The previous experiments which attempted to manipulate recollection all used a response deadline manipulation, controlling the time allowed for recognition decisions. While these paradigms did provide evidence in favour of the theory they prevented conclusive analysis of response latency data, since the manipulation altered response latency. In addition, it is possible that the effects observed were effects generally of a response time manipulation and not of recollection specifically. In light of this, in the next series of experiments I aimed to manipulate the availability of recollected evidence without manipulating response deadline, instead using divided attention manipulations. By starting with a task in which recollection would normally be used, and impairing recollection for some participants, I expected to be able to observe the effects of changes in the availability of recollected evidence on the confidence-accuracy relationship. To foreshadow my results, manipulations checks for all of these experiments failed to demonstrate successful manipulation of recollection using divided attention. So, here, I report the rationale, methodology, and results in full, but note that these experiments were not able to provide a test of my key hypothesis regarding the role of recollection in the confidence-accuracy relationship of negative decisions.

Divided attention at encoding is one manipulation that has been used in many previous studies to impair use of recollected evidence (e.g., Gardiner, Ramponi, & Richardson-Klavehn, 1998; Malmberg & Xu, 2007; Verde, 2004). Divided attention manipulations require participants to engage in a second task at the same time as studying or being tested on items in the recognition task. The secondary task used in past experiments varies, ranging for example from tone monitoring (e.g., Gardiner & Parkin, 1990; Palmer, Brewer, & Horry, 2013) to pressing keys when certain items appear on screen (e.g., Skinner & Fernandes, 2008) to random number generation (e.g., Knott & Dewhurst, 2007). A common finding across various divided attention manipulations is that divided attention selectively impairs recollection, but does not impair or has a smaller effect on familiarity (e.g., Jones & Jacoby, 2001; Skinner & Fernandes, 2009; Yonelinas, 2001). For example, Reinitz, Morrissey, and Demb (1994) had participants perform a dot-counting task as a divided attention manipulation during a face recognition task. Participants were then asked to discriminate between studied faces and new faces made up of features from studied faces. Participants made more false alarms to new faces with familiar features in the divided attention than in the full attention condition. In addition in the divided attention condition participants made more false know ratings on a remember/know test the more familiar features the faces contained, but rates of false remember ratings did not change, suggesting that participants were influenced more by familiarity in the divided attention condition. In an associative recognition task using word pairs, Odegard, Koen, and Gama (2008) divided attention by asking participants to press a button when they heard three consecutive odd numbers during study. Participants made more false alarms to rearranged word pairs in the divided attention condition than the full attention condition. This was taken to indicate that participants' use of recollection was reduced in the divided attention condition compared with the full attention condition as rearranged pairs, containing two studied words, would be difficult to distinguish using familiarity.

In Experiments 6a to e I attempted to impair recollection through use of various divided attention manipulations. Experiment 6a used a variation on a divided attention task used by Otgaar, Peters, and Howe (2012) for children and young adults in which a different coloured smiley appeared before and after each study item, and participants were asked to report how many smileys had been red following the study phase. Experiment 6a used coloured circles and asked participants to report the number of each colour seen. In Otgaar et al. (2012) et al.'s experiment, divided attention led to a reduction in correct recall of studied words, which should mean that recollection is difficult as details from study cannot be retrieved. Experiments 6b to e used digit and sum monitoring tasks to divide attention. These were similar to those used by Skinner and Fernandes (2008), with participants pressing keys to indicate whether the numbers met certain criteria. Skinner and Fernandes found that participants under similar divided attention conditions made more errors in recall of words than those in the full attention condition, thus suggesting that recollection should be impaired by this task.

Audio divided attention tasks such as tone monitoring and audio digit monitoring tasks were not chosen because I hoped to find a divided attention manipulation that could be used during study with the task from Experiment 5, where some of the studied stimuli were presented in audio form. The audio versus text task was used for Experiments 6a b and c and a plurality discrimination task was used for Experiments 6d and e, as both tasks had already demonstrated use of recollection in the previous experiments.

Method

Participants. A total of 139 participants (80 female, age: 18 to 68 years, *M* = 25.15, *SD* = 9.42) took part in the divided attention experiments. Experiment 6a contained 20 participants (11 female, age: 18 to 32 years, *M* = 20.9, *SD* = 3.49), Experiment 6b 20 (12 female, age: 18 to 38 years, *M* = 23.15, *SD* = 4.96), Experiment 6c 40 (23 female, age: 18 to 40 years, *M* = 22.43, *SD* = 5.19), Experiment 6d 17 (12 female, age: 18 to 40 years, *M* = 23.06, *SD* = 7.15), and Experiment 6e 43 (30 female, age: 18 to 68 years, *M* = 31.57, *SD* = 13.19). All participants took part in return for either course credit or remuneration, and had

English as a first language and normal or corrected to normal vision, as well as normal or corrected to normal hearing for the experiments with auditory stimuli.

Design. Five variations on a divided attention manipulation were attempted. Four added a divided attention manipulation during the study phase and one during the test phase. Each experiment used a 2 (presentations: one, three[9]) × 2 (stimulus type: plural, singular or visual, auditory) × 2 (attention condition: full, divided) mixed design. Repetitions and stimulus type were varied within-subjects. Attention condition was varied between-subjects for Experiment 6a and within-subjects for the remaining experiments. The dependent variables were confidence judgments about each recognition decision, accuracy of recognition decisions, and response latency of decisions.

Materials. Stimuli were 128 nouns with the same properties and from the same source as for the previous experiments. Words were randomly allocated in the same manner as for previous experiments to be half auditory and half visual stimuli for the modality judgment tasks and half singular and half plural for the plurality discrimination tasks. Words were also randomly allocated to number of presentations as in Experiment 5. At test, half of the words were randomly allocated to be paired with the correct cue for the audio versus text task (e.g., *Was this voice?* for a word presented in audio form during study) and half to be presented

[9] In Experiment 5, which used the same tasks as Experiments 6a,b and c, one verus three presentations was shown to be adequate to produce an effect on familiarity. In the divided attention experiments it was not desirable to use a larger number of presentations due to the large number of blocks required to use the divided attention manipulations within-subjects. Participants had to complete multiple study and test trials and as such may have found it difficult to attend to the word lists if the duration of study was extended further by increased presentations. However this may present a possible reason for the non-significant plurality discrimination results as the earlier plurality discrimination experiments required more presentations to show an effect on false alarms.

with the opposite cue. For the plurality discrimination tasks half of the words were randomly allocated to be presented at test in the form studied, and half to have changed plurality.

Procedure. Each experiment was broken into eight blocks each containing a study and test phase. Participants studied 20 words in each block. Participants moved through the tasks and instructions in the same manner as for previous experiments. The modality judgment tasks functioned in the same way as in Experiment 5, except that no new words appeared in the test phase. Plurality discrimination tasks were the same as in Experiment 4.

In addition, participants were randomly allocated to either the divided attention or full attention condition, for Experiment 6a, or to complete either the divided attention or full attention condition first, for the remaining experiments. In all of the modality judgment tasks the divided attention manipulation occurred during the study phase. In Experiment 6a, a coloured circle appeared before and after each study word. There were four different circle colours, blue, yellow, pink and red. The experimental software randomised which colour appeared after each word. Participants in the divided attention condition were told that they should try to keep count of the number of each colour that appeared, as they would also be tested on this in the test phase. After the study phase was complete, participants in the divided attention condition were asked to report the number of each colour circle that they had counted by typing it into a text box which appeared on screen. Participants in the full attention condition also saw the coloured circles before and after each study word but were told to ignore them and moved straight on to the test phase after study.

In Experiment 6b, participants saw a number between 1 and 9 before and after each study word. Again the experimental software randomised which number appeared following each study item. In the divided attention condition participants were asked to respond to these numbers by pressing keys to indicate whether each number that appeared was higher or lower than the previous number. In the full attention condition participants were asked to ignore the numbers and concentrate on the words.

In Experiment 6c, participants saw a simple sum of two numbers between 0 and 9 (e.g., 1 + 6) appear before and after each study word. In the divided attention condition they were asked to respond to these sums by pressing keys to indicate whether each total was higher or lower than the previous total. In the full attention condition the sums still appeared, but participants were told to ignore them and focus on the words.

Experiments 6d and 6e used the plurality discrimination task. These tasks were the same as Experiment 4. In Experiment 6d the same divided attention task using sums was used as for Experiment 6c. In Experiment 6e, the sums were again used but this time they appeared during the test phase. Rather than a sum appearing before and after each study item, a sum appeared before and after each test trial, such that participants had to remember the previous sum during each test trial in order to indicate whether the next total was larger or smaller. This variation was used because the response deadline manipulation had affected participants only at test, and therefore I expected divided attention at test to produce more comparable results. In the full attention condition the sums still appeared but participants were asked to ignore them and concentrate on the word test.

Results

To test whether dividing attention successfully impaired recollection, a mixed effects model was created for each experiment and for the combined data from all divided attention experiments with false alarms as the outcome and attention condition and number of presentations as fixed effects. A similar model was also created with hits as the outcome. Table 15 summarises the fixed effects coefficients for each of these models. If dividing attention impaired recollection, I would expect to see false alarms increasing with number of presentations more or only in the divided attention conditions. This would be demonstrated by a significant negative regression coefficient for the interaction between number of presentations and attention condition. As the divided attention condition was coded as zero and the full attention condition as one, a negative regression coefficient for the interaction between presentations and condition would indicate that the effect of presentations on false alarms was weaker in the full attention condition. Hence this would show that divided attention led to a greater increase in false alarms with presentations than full attention.

If hits are predominantly affected by familiarity as suggested by Reder, Angstadt, Cary, Erikson, and Ayers (2000), I would expect hits to increase with presentations in both attention conditions. Therefore I did not expect to find a significant interaction between number of presentations and condition for hits. Another pattern of results would be possible, however. If increased use of recollection allowed participants to correctly identify more studied stimuli, hits could increase more with increasing presentations in the full attention condition. This would be shown by a significant positive regression coefficient for the interaction between attention condition and number of presentations for hits, which would indicate that presentations had a stronger positive effect on hits in the full attention condition.

In all experiments except Experiment 6c, the interaction between number of presentations and attention condition for false alarms was not significant, as shown by the confidence intervals which all include zero. In Experiment 6c there was a significant interaction between attention condition and number of presentations, however the value of *b* was positive, indicating that false alarms increased more with increasing presentations in the full attention condition than the divided attention condition; the opposite of what was expected. In addition, the positive *b* values for the effect of number of presentations in four out of five experiments and overall indicate that in the divided attention condition (here coded as zero) the trend was towards a fewer false alarms following three presentations than following one presentation. However this effect was not significant in any of the experiments or in the overall data. Overall, this result suggests that recollection was not significantly impaired by divided attention in any of these experiments.

In addition the only analyses which showed significant interactions between number of presentations and condition for hits were Experiment 6e and the overall data. The non-

Table 15

Fixed Effects Coefficients for Mixed Effects Models Predicting False Alarms and Hits in each of the Divided Attention Experiments from Attention Condition and Number of Presentations

Fixed Effect	*b*	SE_b	95% CI_b
	Expt 6a, False alarms		
Intercept	-1.20	0.19	[-1.45, -0.82]
Condition, full attention (C)	0.36	0.27	[0.08, 0.74]
Presentations, 3 (P)	-0.63	0.30	[-1.24, 0.01]
C × P	0.60	0.39	[-0.08, 0.86]
	Expt 6a, Hits		
Intercept	0.88	0.12	[1.40, 1.87]
C	-0.35	0.17	[-0.82, -0.29]
P	0.61	0.19	[0.95, 1.65]
C × P	0.18	0.26	[-0.40, 0.76]
	Expt 6b, False alarms		
Intercept	-0.96	0.22	[-2.04, -1.51]
C	-0.17	0.29	[-0.87, -0.08]
P	-0.40	0.19	[-1.25, -0.22]
C × P	-0.10	0.28	[-0.89, 0.42]
	Expt 6b, Hits		
Intercept	1.01	0.20	[0.79, 1.50]
C	-0.01	0.35	[-0.83, 0.78]
P	1.03	0.21	[0.66, 1.35]
C × P	-0.15	0.30	[-0.78, 0.45]

	Expt 6c, False alarms		
Intercept	-1.06	0.21	[-2.67, -1.62]
C	-0.02	0.32	[-0.56, 0.52]
P	-0.09	0.13	[-0.29, 0.07]
C× P	0.26	0.19	[0.09, 0.78]
	Expt 6c, Hits		
Intercept	1.14	0.15	[2.01, 2.59]
C	0.10	0.23	[-0.95, 0.96]
P	0.59	0.14	[0.93, 1.44]
C × P	0.43	0.21	[0.48, 1.26]
	Expt 6d, False alarms		
Intercept	1.42	0.42	[-3.55, -1.84]
C	-0.10	0.56	[-1.17, 1.08]
P	0.13	0.20	[-0.17, 0.63]
C × P	-0.18	0.30	[-0.67, 0.07]
	Expt 6d, Hits		
Intercept	1.56	0.55	[2.16, 3.62]
C	-0.21	0.65	[-1.17, 0.50]
P	0.73	0.23	[1.16, 2.02]
C × P	-0.05	0.34	[-0.84, 0.56]
	Expt 6e, False alarms		
Intercept	-0.76	0.26	[-1.29, -0.16]
C	-0.45	0.31	[-1.13, 0.21]
P	-0.26	0.13	[-0.47, 0.03]
C × P	-0.09	0.19	[-0.43, 0.03]
	Expt 6e, Hits		
Intercept	1.39	0.16	[1.14, 1.63]
C	0.12	0.26	[-0.45, 0.42]
P	0.47	0.15	[0.06, 0.79]
C × P	0.53	0.22	[0.08, 1.03]
	Overall, False alarms		
Intercept	-1.05	0.14	[-2.43, -1.75]
C	-0.06	0.20	[-0.68, 0.34]
P	-0.20	0.07	[-0.55, -0.27]
C × P	0.06	0.11	[-0.05, 0.33]

Overall Hits			
Intercept	1.18	0.11	[0.93, 1.37]
C	-0.03	0.16	[-0.40, 0.42]
P	0.64	0.08	[0.51, 0.76]
C × P	0.28	0.11	[0.10, 0.50]

Note. The within-subjects models (all except Experiment 6a) and the overall model included test item (*SD* = 0.22) and participant (*SD* = 0.66) as random slopes and intercepts. In addition, the overall model included experiment (*SD* = 10) as a random slope and intercept. The model for Experiment 6a included test item (*SD* < 0.01) as a random slope and intercept.

significant interactions between attention condition and number of presentations for the majority of the experiments indicate that that there was no significant difference in participants' ability to identify studied items in the divided attention condition compared with the full attention condition. For Experiment 6e and the combined data the significant interaction indicates that hits increased more with more presentations in the full attention condition than the divided attention condition.

The significant interaction for hits between presentations and attention condition for the combined data may indicate some increased use of recollection, as false alarms did not increase. However changes in hits with presentations did not differ in a consistent way between the conditions across experiments. In three out of five experiments hits increased more with presentations in the full attention condition, but in Experiment 6d there was no significant difference between the conditions in the effect of presentations on hits, and in Experiment 6a the pattern was in the reverse direction with fewer hits with increasing presentations in the full attention condition.

The significant positive *b* values in all experiments for the effect of presentations on hits demonstrate that under divided attention conditions, hits increased with number of presentations in all experiments. However, as no significant interaction was evident in any of the experiments taken alone, none of the experiments individually demonstrated that this effect changed depending on attention condition. In addition it is notable that in all but one of the experiments the regression coefficient for the effect of presentations on false alarms is negative, indicating that three presentations led to fewer false alarms than one presentation in the divided attention condition, although this effect was not significant in all of the experiments. In the combined data, the effect of presentations on false alarms was significant, suggesting that in the divided attention experiments combined, three presentations led to fewer false alarms than a single presentation in the divided attention condition. This is consistent with recollection occurring in the divided attention condition. As the interaction between attention condition and presentations was not significant in the combined data, this suggests that recollection was used similarly in these experiments under divided and full attention.

The patterns of false alarms and hits thus gave no evidence of a difference in recollection between the divided and full attention conditions. Hence conclusions are limited, and it is likely that only a very small effect if any was present. These results suggest that the divided attention manipulations were not successful in impairing recollection to a meaningful extent. The significant negative interaction between number of presentations and attention condition on hits in the combined data may indicate that participants had access to slightly more recollected information in the full attention condition leading to an increased ability to identify studied items. However no conclusive evidence was present to indicate that recollection changed between the conditions.

To test whether changes in any of the outcome variables were likely to be due to a change in response bias I also created a mixed effect model for the combined data from all the experiments with response as the outcome and correct response, presentations and condition as fixed effects. The fixed effects coefficients for this model are shown in Table 16.

Table 16

Fixed Effects Coefficients for the Mixed Effects Model Predicting Response from Correct Answer, Condition and Number of Presentations

Fixed Effect, level	*b*	SE_b	95% CI
Intercept	-1.08	0.09	[-1.26, -1.00]
Answer, negative (A)	2.01	0.07	[1.90, 2.15]
Condition, full attention (C)	0.04	0.10	[-0.07, 0.22]
Presentations, three (P)	-0.62	0.08	[-0.76, -0.44]
A × C	0.12	0.10	[-0.11, 0.33]
A × P	0.79	0.10	[0.61, 0.86]
C × P	-0.25	0.11	[-0.41, -0.08]
A × P × C	0.20	0.15	[0.02, 0.39]

Note. The model included test word (*SD* = 0.18), participant (*SD* = 0.33) and experiment (*SD* = 0.14) as random slopes and intercepts.

The significant positive regression coefficient for correct answer indicates that participants were more likely to give a negative answer when the correct answer was negative, in the divided attention condition following one presentation. This shows that participants were able to make correct rejections when appropriate even under divided attention conditions when they had only studied the word once. There was no significant effect of condition on response. The significant negative b value for the effect of presentations demonstrates that when the correct answer was positive, in the divided attention condition, participants were less likely to give a negative response; that is, they were more likely to give a correct positive response following three presentations than one presentation.

The interaction between correct answer and attention condition was not significant. This indicates that the likelihood of participants making a correct negative decision following one presentation did not significantly change depending on attention condition. The significant positive regression coefficient for the correct answer by presentations interaction shows that participants were more likely to correctly make a negative response following three presentations than following one presentation in the divided attention condition. The significant negative b value for the interaction between attention condition and presentations demonstrates that the negative effect of increased presentations on the likelihood of participants making an incorrect negative decision was stronger in the full attention condition. That is, participants were more likely to correctly identify studied items following three presentations, and this effect was stronger in the full attention condition. This reflects the significant interaction between presentations and conditions on hits shown in the previous analysis.

Finally, the significant positive regression coefficient for the three way interaction between correct answer, condition and presentations suggests that in the full attention condition compared with the divided attention condition, the positive effect of increased presentations on correct rejections was stronger. These results suggest that participants were able to better distinguish between studied and unstudied items following a greater number of presentations and that this effect was greater under full attention conditions. This is consistent with more recollection occurring in the full attention condition, in contrast to the earlier analyses. A final test of the extent to which the attention conditions influenced recollection is the overall effects of attention on false alarms and hits. If recollection was impaired in the divided attention conditions, more false alarms would be expected to occur in the divided attention conditions than the full attention conditions overall, regardless of number of presentations. This was tested by a mixed effects model on the combined data with false alarms as the outcome and only attention condition as a fixed effect, and a similar model with hits as the outcome. The fixed effects coefficients for these models are shown in Table 17.

Table 17

Fixed Effects Coefficients for the Mixed Effects Model Predicting False Alarms and Hits from Attention Condition

Fixed Effect, level	*b*	SE_b	95% CI
	False Alarms		
Intercept	-1.15	0.14	[-2.65, -1.93]
Condition	-0.03	0.19	[-0.46, 0.42]
	Hits		
Intercept	1.48	0.09	[2.77, 3.30]
Condition	0.06	0.17	[-0.41, 0.46]

Note. The models included test item (*SD* = 0.30) participant (*SD* = 0.93) and experiment (*SD* = 0.12) as random slopes and intercepts.

Attention condition did not have a significant effect on either false alarms or hits, as shown by the 95% confidence intervals for both which do not include zero. This suggests that attention did not influence recollection, and the significant effects shown in the previous analysis may have been due to changes in response bias produced by the presentations manipulation. Hence there was no evidence of divided attention significantly impairing recollection.

Assuming that the lack of evidence for a difference in recollection means that recollection did not change with the divided attention manipulation and was used to a large extent in both attention conditions, my theory would suggest that if a positive-negative was present, it should be very small reflecting the large amount of recollection used, or even not present, and that this should not change depending on attention condition. To test whether this was the case, I created a mixed effects model using the overall combined data from all divided attention experiments, with confidence, response type and attention condition as fixed effects, and accuracy as the outcome. The fixed effects coefficients for these models are shown in Table 18. A positive-negative difference in the typically observed direction would be demonstrated if a significant negative regression coefficient for the interaction between confidence and response was present. A difference in the size or direction of the positive-negative difference depending on attention condition would be shown by a significant interaction between condition, response, and confidence. If this interaction was associated with a negative *b* value, it would indicate that the positive-negative difference was smaller in the full attention condition. If it was positive, it would indicate that the positive-negative difference was smaller in the divided attention condition. If as the data on false alarms suggests, recollection did not differ between the attention conditions, I would not expect to see a significant interaction between condition, confidence and response. If there was a small

Table 18

Fixed Effects Coefficients for the Mixed Effects Models Predicting Accuracy from Confidence, Attention Condition and Response and Response Latency, Attention Condition and Response in the Divided Attention Experiments Overall

Fixed Effect	*b*	SE_b	95% CI_b
	Confidence		
Condition, divided (C)	0.47	0.25	-0.09, 1.22
Confidence (conf)	0.04	0.002	0.03, 0.04
Response, negative (R)	0.41	0.24	-0.06 0.87
C × Conf	-0.01	< 0.01	-0.02, < -0.01
C × R	0.83	0.34	-1.62, -0.24
Conf × R	< -0.01	<0.01	0.01, < 0.01
C × R × Conf	0.01	< 0.01	< 0.01, 0.02
	Response Latency		
C	-0.31	0.07	[-0.48, -0.19]
Response latency (RL)	-0.92	0.16	[-1.36, -0.55]
R	-0.08	0.08	[-0.20, 0.05]
C × RL	0.79	0.19	[0.42, 1.28]
C × resp	0.31	0.11	[0.06, 0.47]
R × RL	0.57	0.22	[0.15, 1.09]
C × R × RL	-1.12	0.30	[-1.85, -0.35]

Note. The models included test item (SD =), participant (SD =) and experiment (SD =) as random slopes and intercepts.

difference in recollection as the pattern of hits suggested was possible, a small interaction might occur. However this would not provide conclusive evidence for my theory as the extent to which participants used recollection in either condition is not clear, and therefore it is not possible to determine whether the confidence-accuracy relationship was affected by recollection or some other variable.

In the divided attention condition (coded as zero), no positive-negative difference was observed for confidence, as shown by the confidence interval for the interaction between response and confidence which includes zero. In addition the regression coefficient for this interaction is close to zero. For response latency, a positive *b* value indicated that negative decisions displayed a weaker negative response latency-accuracy relationship than positive decisions, demonstrating a positive-negative difference in the direction normally observed. This effect was also significant as shown by the confidence interval not including zero.

For the confidence-accuracy relationship the attention conditions did not differ significantly, indicating that attention did not significantly change the positive-negative difference. Since there was no positive-negative difference present in the divided attention condition, as indicated by the non-significant regression coefficient for the response by confidence interaction, this indicates that there was no significant positive-negative difference for the confidence-accuracy relationship in either condition. For the response latency-accuracy relationship, the positive-negative difference was significantly smaller in the full attention condition, as shown by the significant positive *b* value for the three way interaction between response latency, response and condition.

Figure 9 shows this interaction with surprising differences between conditions. The effect of dividing attention on the response latency-accuracy relationship was actually reversed for positive versus negative decisions. For positive decisions, a stronger relationship was evident in the full attention condition. However, for negative decisions, the reverse effect was evident, with a stronger relationship between response latency and accuracy in the divided attention condition.

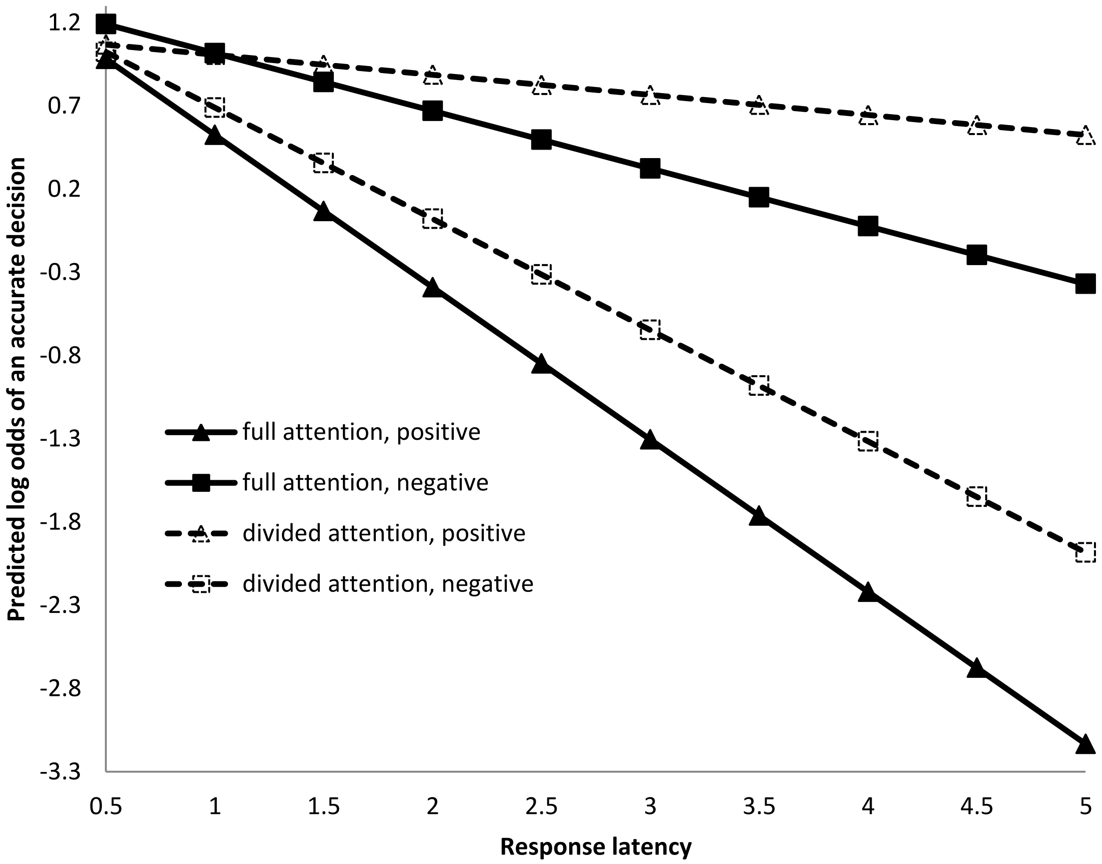

Figure 9. The Response latency-accuracy relationship for positive versus negative decisions in the full versus divided attention conditions in the divided attention experiments overall.

Discussion

Overall the various divided attention manipulations did not appear to impair recollection in these experiments. Therefore, the conclusions that can be made from the confidence-accuracy and response-latency-accuracy data are limited. The confidence-accuracy results were consistent with what my theory predicts if recollected evidence were highly available in both conditions.

A question which must be considered here is why divided attention did not appear to impair recollection in these experiments when it has been established as doing so in past experiments in the literature. One possibility involves the way in which recollection was quantified in these experiments compared with previous work. In many previous studies with divided attention recollection has been measured using remember/know judgments and quantified as an all or nothing variable (e.g., Palmer et al., 2013). However recollection has been shown to occur in a continuous fashion, and to contribute in varying ways to different aspects of a recognition decision (e.g., Wixted & Mickes, 2010). In addition, remember judgments have often been shown to reflect stronger memories than know judgments. Thus a remember judgment may only reflect a specific strong type of recollection, which may be reduced to a greater extent by divided attention than more moderate or small quantities of recollected detail. Other studies have noted effects of dividing attention on the amount of erroneous recollected information rather than a reduction in recollected information produced (e.g., Skinner & Fernandes, 2008) or in general used self-report ratings of recollection which may differ from the objective measures used in these experiments (e.g., Palmer, Brewer, McKinnon, & Weber, 2010). A useful addition to future studies may be the inclusion of a subjective continuous measure of recollection as well as the kind of measures used here. However, based on the false alarm and hit rates observed, it appears that recollection at least of the type measured by these was unaffected by divided attention. The results concerning the confidence-accuracy relationship are consistent with what the proposed explanation of the positive-negative difference would suggest should occur if recollection occurred in both conditions, as will be discussed below.

No significant positive-negative difference was present in the confidence-accuracy relationship for the divided attention condition, and full attention did not significantly change the positive-negative difference, as shown by the confidence interval for the three way interaction between confidence, response and condition, which includes zero. Thus the results are consistent with a high level of recollection occurring in both conditions. This provides some limited support for the proposed theory as where there was evidence that levels of recollection were high, no positive-negative difference was present.

The response latency-accuracy relationship demonstrated a positive-negative difference, however this was affected by attention condition an unusual way, with negative decisions showing a stronger relationship under divided attention conditions while positive decisions demonstrated a stronger relationship under full attention conditions. Again this seems to reflect different mechanisms underlying the time course of positive and negative decisions. However, there was no evidence that this was caused by differences in recollection. As the response latency-accuracy relationship is not the focus of this thesis, suggestions for what may have caused the variation in the response latency-accuracy relationship are not proposed here. However, the results do suggest that divided attention can have effects on the response latency-accuracy relationship which are independent of changes in recollection.

Since none of the divided attention manipulations showed clear evidence of impairing recollection, the logical next step was to seek a different means of manipulating the extent to which recollection could be used in an associative recognition task, and determine whether the results found with the response deadline task would replicate with another type of manipulation. Rather than seeking to impair recollection, finding a manipulation that would enhance participants’ ability to use recollected evidence would have more practical relevance since a successful manipulation of this type would suggest possibilities for improving the confidence-accuracy relationship for negative decisions. Chapter 7 describes an experiment to investigate this possibility.

References

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*(6), 579-583. doi: 10.3758/bf03197100

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7*(1), 1-26. doi: 10.1006/ccog.1997.0321

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*(1), 82-102. doi: 10.1006/jmla.2000.2761

Knott, L. M., & Dewhurst, S. A. (2007). The effects of divided attention at study and test on false recognition: A comparison of DRM and categorized fists. *Memory & Cognition, 35*(8), 1954-1965. doi: 10.3758/bf03192928

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*(3), 545-556. doi: 10.3758/bf03193293

Odegard, T. N., Koen, J. D., & Gama, J. M. (2008). Process demands of rejection mechanisms of recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(5), 1296-1304. doi: 10.1037/a0013034

Otgaar, H., Peters, M., & Howe, M. L. (2012). Dividing attention lowers children's but increases adults' false memories. *Journal of Experimental Psychology-Learning Memory and Cognition, 38*(1), 204-210. doi: 10.1037/a0025160

Palmer, M. A., Brewer, N., & Horry, R. (2013). Understanding gender bias in face recognition: Effects of divided attention at encoding. *Acta Psychologica, 142*(3), 362-369. doi: 10.1016/j.actpsy.2013.01.009

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica, 133*(2), 137-145. doi: 10.1016/j.actpsy.2009.11.002

Reder, L.M., Angstadt, P., Cary, M., Erikson, M.A. , & Ayers, M.A. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Learning, memory, and cognition, 26*, 294-320. doi: 10.1037/0278-7393.26.2.294

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(1), 161-168. doi: 10.1037/0278-7393.20.1.161

Skinner, E. I., & Fernandes, M. A. (2008). Interfering with remembering and knowing: Effects of divided attention at retrieval. *Acta Psychologica, 127*(2), 211-221. doi: 10.1016/j.actpsy.2007.05.001

Skinner, E. I., & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults under divided attention. *Psychology and Aging, 24*(1), 211-216. doi: 10.1037/a0014177

Verde, Michael F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition (pre-2011), 32*(8), 1265-1272.

Wixted, J. T., & Mickes, Laura. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025-1054. doi: 10.1037/a0020874

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology-General, 130*(3), 361-379. doi: 10.1037//0096-3445.130.3.361

CHAPTER 7 — INCREASING USE OF RECOLLECTION

The experiments in this thesis aimed to test the idea that the weak relationship between confidence and accuracy for negative recognition decisions is due to neglect of recollected evidence. If this is the case, impairing individuals' ability to use recollected evidence should weaken the confidence-accuracy relationship for negative decisions while improving the degree to which individuals use this type of evidence should strengthen the relationship. The previous experiments demonstrated that impairing individuals' ability to use recollected evidence through enforcing a response deadline during study weakens the relationships between accuracy and its predictor variables for negative decisions. This supports the idea that neglect of recollected evidence when making negative decisions may be responsible for the weak relationship between confidence and accuracy for these decisions. However it does not tell us whether anything can be done to enhance the confidence-accuracy relationship based on this information. The aim of Experiment 7 was to investigate whether increasing the availability of recollected evidence could strengthen the relationship between confidence and accuracy. The extent to which knowing about the memorial bases of recognition decisions can impact on real world problems depends on a) whether individuals' use of recollection can be increased and b) whether doing so can improve an assessor's ability to use confidence to determine the accuracy of a decision. This would necessarily involve being able to encourage use of recollected evidence and observe strengthening of the relationship between these predictors and accuracy.

One potential method of increasing the availability of recollected evidence is to implement a recall phase before a recognition test, requiring individuals to attempt to recollect the correct response before making a recognition decision about whether or not a suggested answer is correct. By attempting to retrieve the item before attempting the recognition test, individuals will have one of two types of evidence available for their decision and confidence rating. First, they may have the item they recall in mind before the recognition phase, allowing them to easily utilise this recollected evidence to contrast or match the item being tested. Second, they may be aware that they cannot recall the item they studied, reducing the likelihood of a confident negative response based on poor memory, and hence low familiarity. Evidence suggests that prior recall leads to fewer false alarms in a recognition test (Huff, Davis, & Meade, 2013), potentially due only to changes in recall (Chan & McDermott, 2007) and recent work in the Flinders labs indicates that confidence and accuracy are more strongly related for negative decisions about questions to which individuals can recall an answer (Weber, Woodard, Williamson, & Hogan, 2014). Therefore I would expect that following a recall test on studied information, participants would be more capable of using recollected information about the same stimuli during a recognition test.

In Experiment 7, I implemented a recall phase for half of the participants by asking them to study word pairs, and at test supplying one word from a pair and asking them to recall the other, while the other half of participants only completed the recognition test on the word pairs. I expected that for negative decisions there would be an interaction between recall condition and confidence on accuracy, such that participants' confidence in their recognition decisions would be more strongly related to accuracy in the recall condition than in the recognition only condition.

Experiment 7

Method

Participants. Eighty Flinders University students (58 female), ranging in age from 17 to 62 years (*M* = 24.49, *SD* = 8.86) participated in return for either course credit or remuneration. All were required to have English as a first language and normal or corrected-to-normal vision.

Materials. Stimuli were 256 nouns between 4 and 10 letters in length, from the same source and using the same ranges of imagability, concreteness and written frequency as for previous experiments. All assignments of words to conditions were done randomly for each participant. The words were randomly allocated into pairs, study blocks, to be presented twice or five times, and to be presented intact or rearranged at test.

Procedure. Participants completed the experiment on computers in individual rooms after reading a brief description of the experiment and giving informed consent. The experiment was broken into four study-test blocks, with participants studying 20 word pairs in the study phase of each block, with half of the words presented twice and the other half five times. Participants were randomly assigned to either a recall-first condition or a recognition-only condition. In the recollection-first condition, participants completed a recall test prior to the recognition test for each word. For the recall test, participants were told: *On each trial you will see a word with a text box below it. Try to remember the pair word that you studied with the word shown. If you can remember the pair word, type it into the text box. If you cannot remember the pair word, type a question mark. It is better to type a question mark than to guess, but please try your best to think of the pair word.*During each trial one word from a pair was supplied and a cue was presented asking *what was presented with….*Below the word was a prompt in smaller font and in brackets saying *[Type ? if you don't know].*

After responding to the recall test participants completed a recognition test containing either an intact studied pair of words containing the cue word from the recall test, or a rearranged pair containing the cue word and a word from another studied pair. In the recognition test, a word pair appeared on screen with the question *were these words paired together?* At the beginning of the test phase it was explained to participants that word pairs may be rearranged, and they should only respond *yes* to pairs that were exactly as they had been during study. An example study pair was given with intact and rearranged versions that could be presented at test. In the recognition test participants were required to click on either the *yes* or the *no* button which were displayed below the words. Following their recognition decision, all participants rated their confidence that their recognition decision was correct by clicking on a button in 10% increments from 50-100%. Participants' response latency for both the recognition test and the confidence response was recorded. In the recognition-only condition, participants did not complete the recall test but proceeded with the recognition trials and confidence ratings only during the test phase.

Results

Because the manipulation in this experiment was intended to enhance recollection, I did not expect an increase in false alarms with number of presentations in the recognition-only condition, since the levels of recollection used in this condition could still be high. However, increased use of recollected evidence in the recall condition could be visible in the data on false alarms and hits in the form of either increased hits in relation to false alarms in the recall condition, or decreased false alarms in the recall condition. To investigate either false alarms or hits differed between the recall-first condition and the recognition-only condition, I created a mixed effects model with false alarms as the outcome and test condition and number of presentations as fixed effects. I then created a similar model with hits as the outcome in the place of false alarms. The fixed effects coefficients for these models are shown in Table 22. For false alarms, the regression coefficient for presentations is positive, suggesting that when presentations increased from two to five, the predicted log odds of a false alarm increased in the recognition only condition. As the 95% confidence interval includes zero, this effect was significant. For the fixed effect of condition, the *b* value is negative suggesting a trend in the direction of fewer false alarms in the recall condition than the recognition only condition

Table 22

Fixed Effects Coefficients for Mixed Effects Models Predicting False Alarms and Hits from Condition and Number of Presentations in Experiment 7

Fixed Effect, level	*b*	SE_b	*95% CI*
	False alarms		
Intercept	-2.34	0.17	[-2.55, -1.98]
Presentations, 5 (P)	0.56	0.21	[0.09, 0.88]
Condition, recall-first (Cond)	-0.13	0.24	[-0.63, 0.12]
Cond × P	-0.26	0.31	[-0.67, 0.27]
	Hits		
Intercept	0.82	0.10	[0.18, 1.46]
P	1.16	0.18	[0.72, 1.48]
Cond	0.09	0.43	[-0.67, 1.01]
Cond × P	-0.07	0.25	-0.51, 0.48]

Note. The models included test word (*SD* = 0.17) as a random slope and intercept.

following 2 presentations. However, as the 95% confidence interval includes zero, this effect was not significant. For the interaction, the regression coefficient is negative, However, this effect was also non-significant. This indicates that when presentations increased from two to five, the effect of condition did not significantly change

For hits, in the recognition only condition, the *b* value for the effect of presentations is also positive, indicating that, like false alarms, the predicted log odds of a hit was higher following five presentations than following two presentations. As the 95% confidence interval does not include zero, this effect was significant. For the fixed effect of condition, the *b* value is positive, indicating that when the number of presentations was two, the recall condition produced slightly more hits than the recognition-only condition, however the 95% confidence interval for this effect includes zero, and hence it was not significant. For the interaction between condition and presentations, the negative *b* value indicates that the effect of increasing presentations may have been slightly smaller in the recall condition, however this effect was also non-significant.

The above analyses did not show significant interactions for either false alarms or hits. Hence there was no clear evidence in the patterns of false alarms and hits to indicate a change in recollection between the recognition-only and recall-first conditions. Overall both hits and false alarms increased with presentations in the recognition-only condition as indicated by the positive *b* values for the effect of presentations. As the interaction between condition and presentations was not significant for either false alarms or hits, the effect of presentations did not differ significantly between the recognition-only and recall conditions. Therefore it appears that familiarity was used to some extent in both conditions, producing increased false alarms. Hits increased more than false alarms with increasing presentations, suggesting that participants were using some recollection in both conditions. This was expected, as the manipulation did not impair recollection, and the task is one in which all test stimuli should be familiar. As all test pairs contained two studied words, familiarity should be relatively similar for targets and foils at test, and therefore participants would have to have used recollection to accurately discriminate between studied and unstudied pairs (Malmberg, 2008; Malmberg & Xu, 2007). Use of recollection in the recognition-only condition could offset the increase in familiarity caused by increased presentations similarly to the recall-first condition. Therefore the fact that there was not a significant difference between the recognition-only condition and the recall-first condition does not indicate that recollection did not increase in the recall-first condition. The fact that hits did not increase is more of a challenge to the prediction that more recollected evidence would be more available in the recall-first condition. Having more recollected evidence available could be expected to add to the evidence supporting positive decisions for studied word pairs, and thus ensure that more of the studied pairs were identified. On the other hand, it is possible that adequate evidence of both kinds was already available in the recognition-only condition.

To further investigate the extent to which participants could discriminate studied and rearranged word pairs in the two conditions, a mixed effects model was also created examining response bias, with response as the outcome and correct answer, condition and presentations as fixed effects. The fixed effects coefficients for this model are displayed in Table 23. The fixed effect of correct answer shows the change in the predicted log odds of

Table 23.

Fixed Effects Coefficients for Mixed Effects Models Predicting Response from Correct Answer, Condition and Number of Presentations

Fixed effect, level	*b*	SE_b	*95% CI*
Intercept	-0.66	0.10	[-0.96, -0.44]
Correct answer, negative (A)	3.00	0.20	[2.61, 3.39]
Condition, recall (C)	-0.05	0.14	[-0.32, 0.18]
Presentations, 6 (P)	-0.99	0.16	[-1.25, -0.60]
C × A	0.19	0.28	[-0.16, 0.62]
C × P	0.06	0.23	[-0.38, 0.39]
A × P	0.43	0.27	[-0.12, 0.82]
C × A × P	0.21	0.39	[-0.20, 0.80]

Note. The models included test word (*SD* = 0.23) as a random slope and intercept.

participants giving a negative response when the correct answer was negative, in the recognition only condition when stimuli had been presented twice. As the *b* value is positive, participants were more likely to give a negative response when a negative response was the correct answer. The 95% confidence interval does not include zero, so this effect was significant. This shows that participants could distinguish rearranged pairs from studied pairs even when they had only studied the target pair twice and only completed the recognition test. That is, participants were able to make correct rejections in the recognition-only condition after only two presentations.

The only other fixed effects for which the regression coefficient was significant was that of presentations. For the fixed effect of presentations, the negative regression coefficient shows that following six presentations compared with two presentations, participants in the recognition-only condition when the correct answer was positive were less likely to give a negative response. That is, following six presentations, participants were less likely to give an incorrect negative response in the recognition-only condition.

None of the interactions between the fixed effects were significant; this is depicted in Figure 12. Hence increased presentations resulted in fewer incorrect negative decisions in both conditions, but there were no significant differences in this effect between the conditions. This is consistent with recollection being used overall, as only recollected evidence could distinguish rearranged pairs from studied pairs containing the same words, and hence sharing all features, given that familiarity is assumed to stem from global match between the test stimulus and all relevant detail in memory (Jones & Jacoby, 2001). It does not, however provide any evidence of increased recollection in the recall condition.

As there is some evidence in the literature to suggest that recollection can occur to different degrees and be used in different ways for different types and parts of a decision

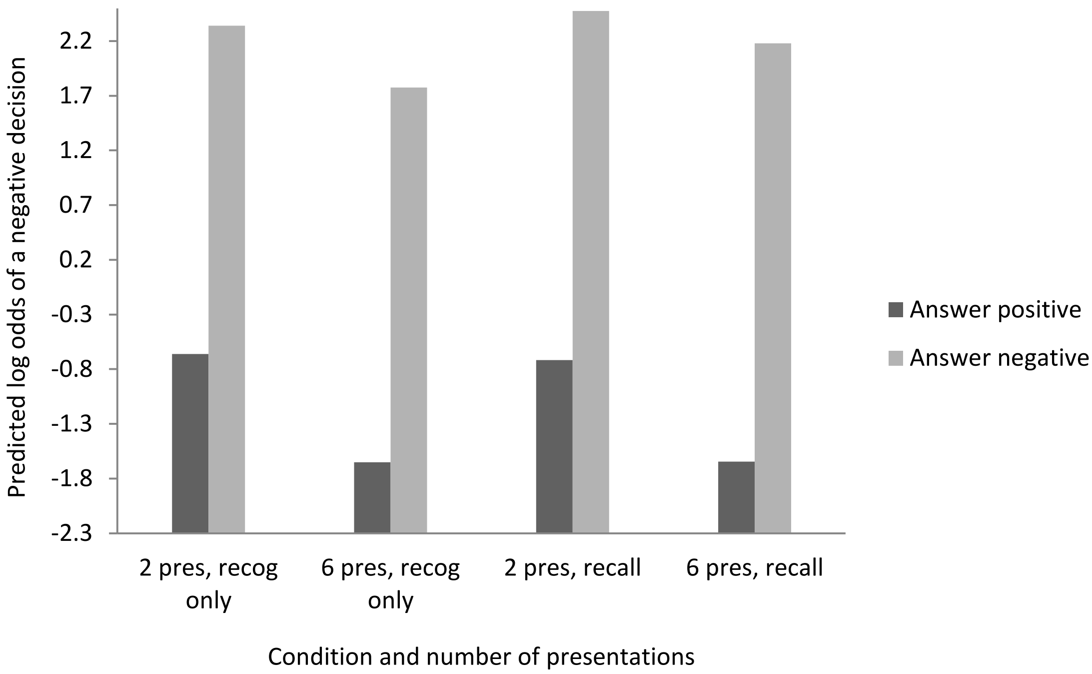

Figure 12 Predicted log odds of a negative decision by correct answer for number of presentations and condition.

(e.g., Ingram, Mickes, & Wixted, 2012; Mickes, Seale-Carlisle, & Wixted, 2013), there is a possibility that recollected evidence could be used to a greater extent in rating confidence without altering recognition responses. This could result in an improved confidence-accuracy relationship for negative decisions due to more detailed evidence being available on which to base confidence ratings. For this reason and two others it was still of interest to analyse the confidence-accuracy data. First, the analysis above, although it did not support increased recollection in the recall condition, also did not suggest that there was any change in response bias between the conditions. Therefore one alternative explanation for any difference in the confidence-accuracy relationship was ruled out. Second, a manipulation which improves the confidence-accuracy relationship for negative decisions is of applied interest even if the mechanism which causes this improvement is different from the one hypothesised. Therefore analyses were conducted to determine whether the manipulation resulted in changes to the confidence-accuracy relationship.

To test the hypothesis that confidence and accuracy would be more strongly related for negative decisions in the recall condition, I created a mixed effects model looking at negative decisions alone with accuracy as the outcome and confidence and test condition as fixed effects. The fixed effects coefficients for this model are shown in Table 24, along with a

Table 24

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy from Confidence and Condition

Fixed Effect, level	*b*	SE_b	95% CI
	Confidence		
Intercept	0.21	0.08	[0.01, 0.34]
Confidence (conf)	0.34	0.02	[0.30, 0.39]
Condition, recall-first (cond)	-0.26	0.10	[-0.44, -0.04]
Cond × Conf	0.11	0.03	[0.08, 0.15]
	Response Latency		
Intercept	1.21	0.07	[1.06, 1.37]
Response latency (RL)	0.02	0.10	[-0.32, 0.31]
Cond	0.14	0.10	[-0.07, 0.31]
Cond × RL	-1.15	0.34	[-1.58, -0.60]

Note. The models included test word (SD = 0.09) as a random slope and intercept.

similar model to investigate the response latency-accuracy relationship. If confidence and accuracy were more strongly related in the recall condition as predicted, I expected to see a significant interaction with a positive regression coefficient between confidence and condition. This would occur if the relationship between confidence and accuracy more positive in the recall-first condition than the recognition-only condition. I expected the relationship between confidence and accuracy to be positive in the recognition-only condition, consistent with the finding across all the previous experiments of a positive confidence-accuracy relationship. However I expected this relationship to be stronger in the recall-first condition.

The regression coefficient for the fixed effect of confidence is positive, as expected, and the 95% confidence interval does not include zero, indicating that this effect is significant. Thus, confidence and accuracy were significantly positively related in the recognition-only condition. The regression coefficient for the interaction between confidence and test condition was positive, indicating that confidence and accuracy were more strongly related in the recall condition, as predicted. This effect was significant, as shown by the confidence interval which does not include zero. Figure 13 depicts this interaction. Thus confidence and accuracy were positively related in the recognition-only condition, but the relationship between confidence and accuracy was stronger in the recall-first condition, as expected.

A similar analysis was conducted for response latency, to determine whether, as expected, it followed similar patterns to confidence. I created mixed effects model with accuracy as the outcome and response latency and condition as fixed effects, for which the fixed effects coefficients are also shown in Table 24. If response latency and accuracy demonstrated a stronger negative relationship in the recall condition, I expected to observe a significant negative regression coefficient for the interaction between response latency and condition. This would occur because response latency would normally be negatively related to accuracy, but would be more strongly (more negatively) related to accuracy in the recall condition. Figure 14 depicts the interaction that occurred. The regression coefficient for the

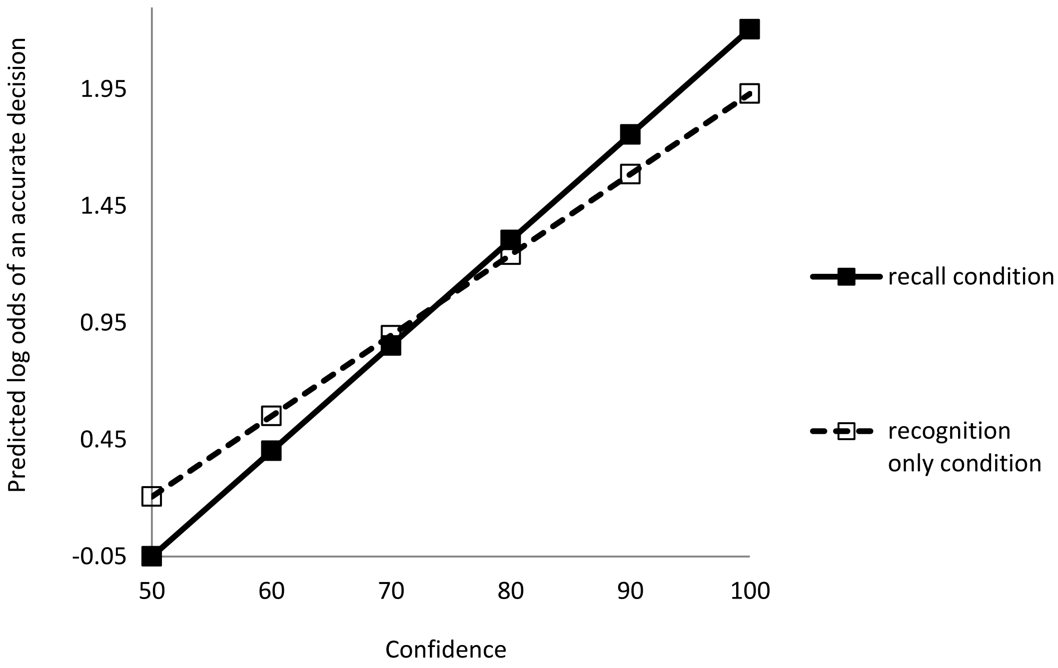

Figure 13. Predicted Log Odds of an Accurate Negative Decision by Confidence and Test Condition.

fixed effect of response latency was not significant. Thus, response latency did not predict accuracy in the recognition-only condition.

The interaction between response latency and condition, however, was significant. as the regression coefficient for this interaction was negative. This demonstrates that the negative relationship between response latency and accuracy was stronger in the recall condition, as expected. This supports my expectation that the response latency-accuracy relationship would follow similar patterns to the confidence-accuracy relationship.

Finally, it was of interest to determine whether the recall phase reduced the positive-negative difference. If this were the case, then the manipulation would show potential for applied usefulness as a method of enabling confidence to indicate accuracy of all decision types. To investigate whether a positive-negative difference was present and whether this

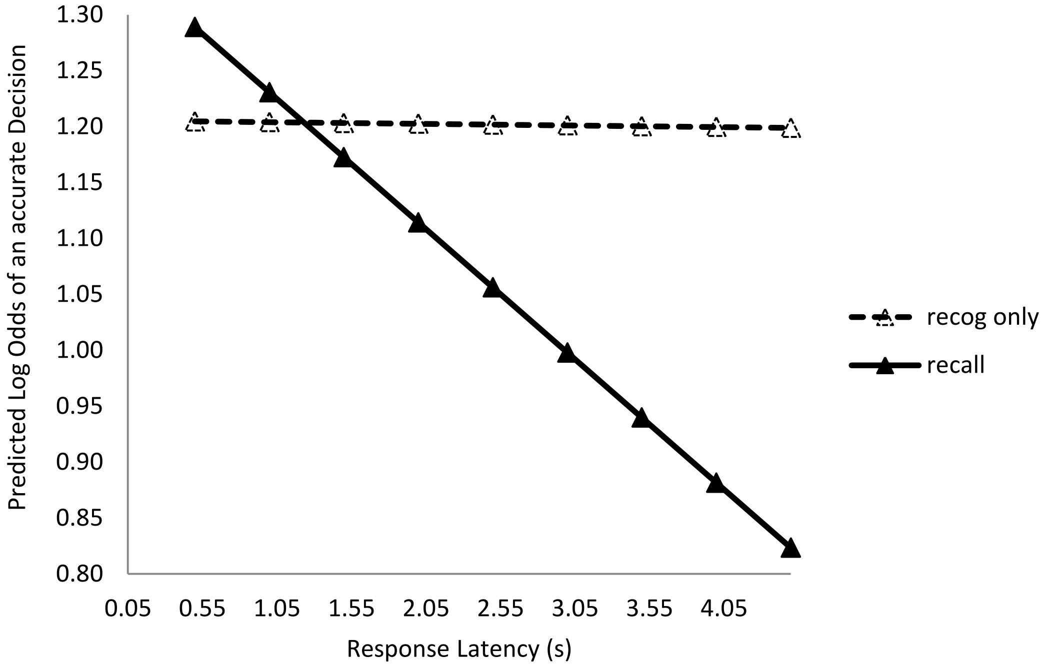

Figure 14. Predicted Log Odds of an Accurate Decision by Response Latency and Test Condition.

changed depending on condition, I created an mixed effects model again with accuracy as the outcome but this time including response, as well as condition and confidence, as fixed effects. The fixed effects coefficients for this model are shown in Table 25. I then created a similar model with response latency in the place of confidence to determine whether theresponse-latency accuracy relationship followed similar patterns to the confidence-accuracy relationship (see Table 25). If a positive-negative difference was present for confidence, this would be shown by a significant negative regression coefficient for the

Table 25.

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy from Confidence, Condition and Response

Fixed Effect, level	*b*	SE_b	*CI Norm*
	Confidence		
Intercept	-0.52	0.14	[-0.75, -0.31]
Confidence (C)	0.62	0.04	[0.58, 0.67]
Response, negative, resp)	0.73	0.16	[0.55, 1.00]
Condition, recall(Cond)	0.16	0.17	[-0.03, 0.48]
Cond × C	0.04	0.05	[-0.05, 0.09]
C × Resp	-0.28	0.04	[-0.35, -0.23]
Cond × Resp	-0.42	0.20	[-0.77, -0.29]
Cond × Resp × C	0.06	0.06	[0.02, 0.17]
	Response Latency		
Intercept	2.28	0.10	[2.09, 2.45]
Response latency (RL)	-0.20	0.03	[-0.24, -0.13]
Resp	-1.07	0.12	[-1.38, -0.74]
Cond	0.13	0.14	[-0.09, 0.43]
RL × Resp	0.20	0.03	[0.10, 0.27]
RL × Cond	-0.14	0.05	[-0.25, -0.05]
Resp × Cond	0.01	0.18	[-0.41, 0.35]
Cond × Resp × RL	0.02	0.06	[-0.06, 0.13]

Note: Models included stimulus word (*SD* = 0.15) as a random slope and intercept.

interaction between confidence and response. This would occur if confidence predicted accuracy for positive decisions, but the relationship between confidence and accuracy was weaker for negative decisions. If the positive-negative difference was then affected by condition, this would be shown by a significant, positive regression coefficient for the three way interaction between confidence, response, and condition. This would indicate that the negative relationship between confidence and response was smaller (due to being more positive) in the recall condition.

The regression coefficient for the three way interaction between confidence, response and condition was positive, and was significant as shown by the 95% confidence interval which does not include zero. This demonstrates that the negative effect of response on the confidence-accuracy relationship was smaller in the recall condition. Thus, the positive-negative difference was smaller in the recall condition. This interaction is represented by Figure 15. The difference in the size of the positive-negative difference can be seen by comparing the slopes of the lines, which indicate the strength of the confidence-accuracy relationship for each condition. The vertical positions of the lines on the graph show the average accuracy of decisions in each condition, but do not describe the confidence-accuracy relationship. Although both decision types were more accurate in the recall condition, the confidence-accuracy relationship for negative decisions improved more in response to the recall manipulation, closing the gap between positive and negative decisions. Hence the positive-negative difference was not eliminated, but was reduced by introducing a recall phase.

For the confidence-accuracy model, the fixed effect of confidence has a positive regression coefficient indicating that the relationship between confidence and accuracy was positive in the recognition-only condition for positive decisions as expected. This effect was significant, as shown by the 95% confidence interval. The confidence by response interaction

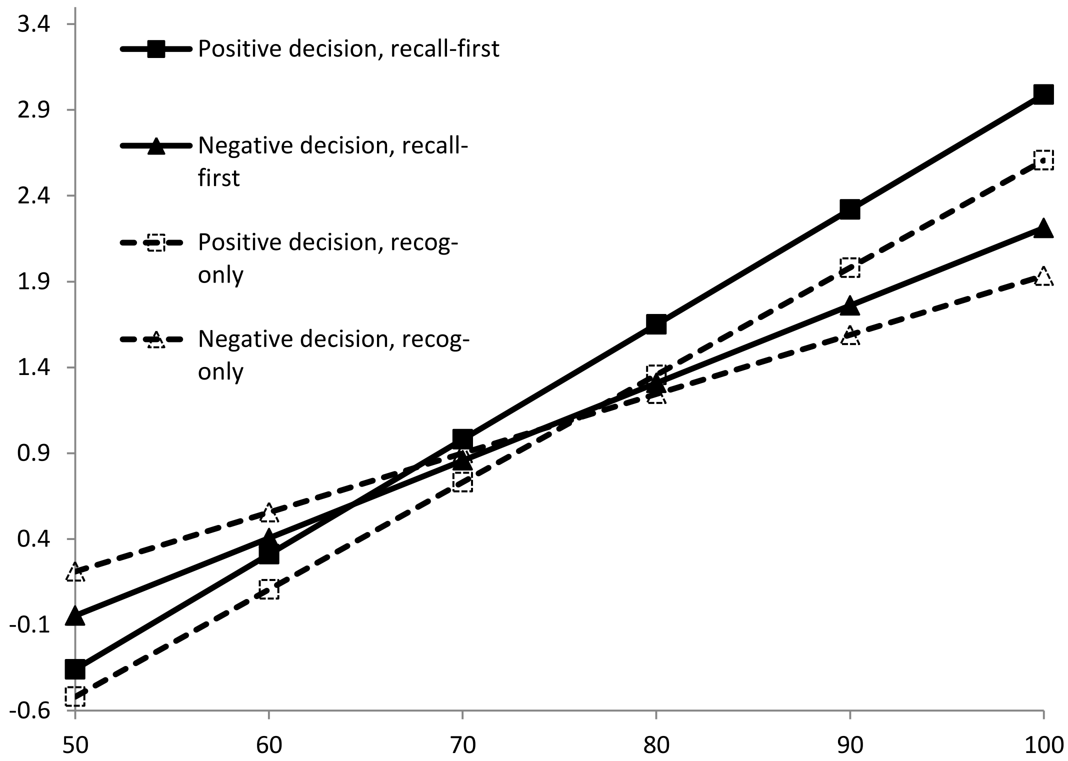

Figure 15. Predicted Log Odds of an Accurate Decision By Decision Type, Condition and Confidence

was negative, as shown by the negative *b* value, and significant, as shown by the 95% confidence interval. This indicates that a positive-negative difference occurred in the recognition-only condition with the positive confidence-accuracy relationship being smaller for negative decisions.

For the response latency-accuracy model, The three way interaction between response type, response latency and condition was not significant, as shown by the confidence interval which includes zero. The negative relationship between response latency and accuracy was stronger in the recall condition for both decision types, but the positive-negative difference was not reduced as the recall condition improved the response latency-accuracy relationship for both positive and negative decisions to a similar degree.

The response latency by response interaction was positive, as shown by the positive regression coefficient, meaning that the negative response latency-accuracy relationship in the recognition-only condition was smaller (more positive, thus, less strongly negative) for negative decisions. This effect was significant. Hence a positive-negative difference was present for response latency in the recognition-only condition, and this difference did not change significantly in the recall-first condition.

Discussion

Overall the results of Experiment 7 supported the idea that the confidence-accuracy relationship for negative decisions can be improved through implementation of a recall phase prior to the recognition test. These findings are promising for practical applications in which the confidence-accuracy relationship for negative decisions is of interest, for example for eyewitness lineup rejections, where the rejection should provide evidence that the suspect is innocent (Wells & Lindsay, 1980). The addition of the recall phase not only strengthened the confidence-accuracy relationship for negative decisions but in doing so reduced the size of the positive-negative difference. The previous experiments in this thesis, despite demonstrating that the confidence-accuracy relationship could be improved for negative decisions, did not find that this made the relationship more comparable with positive decisions. Hence the findings of Experiment 7 are unique in showing that the positive-negative difference can be reduced.

Despite the fact that this experiment suggested a method of improving the confidence-accuracy relationship for negative decisions, however, none of the manipulation checks in this experiment provided conclusive evidence that the effect of the manipulation on the confidence-accuracy relationship was due to increased recollection. Therefore some alternative explanations must be considered for how a recall phase might improve the confidence-accuracy relationship of negative decisions. One possibility is that the recall phase increased participants' awareness of poor memory. Items for which memory is poor would not necessarily produce good recognition decisions, as participants would have very little memorial evidence on which to base their choice. However, awareness that they had failed to recall the pair word just before the recognition test may have led participants to give low confidence ratings to decisions based on poor memory. In contrast, in the recognition-only condition participants may have been less aware that their memory quality was poor and thus taken lack of familiarity for items for which they had poor memory as evidence for a negative decision. This could have led them to give higher confidence ratings to these poor memory decisions in the recognition-only condition than the recall-first condition.

In addition, previous evidence suggests that prior recall predominantly enhances individuals' ability to bring to mind the specific stimuli recalled, and may even reduce the accessibility of other items which share similar features in memory which are not the recalled stimulus (e.g., Chan & McDermott, 2007; Verde, 2004). If accessibility of similar items in memory was reduced by the recall phase in this experiment, participants may have had less access to misleading familiarity evidence stemming from pre-experimental familiarity with certain words and presentation of similar items. This could lead to memorial evidence used in confidence ratings for the recall-first condition being more relevant to the specific item studied

As well as suggesting a possible reason for more realistic confidence ratings, the fact that recalling an item should make it easier to recall later would suggest that adding a recall phase should only provide more ease of producing recollected information for the specific item recalled, not simply increase overall memory quality. However, the overall increase in accuracy in the recall phase in Experiment 7 suggests that a change in memory quality (enhancement of both familiarity and recollection) could be responsible for the changes in the confidence-accuracy relationship. Increased memory quality can improve individuals' ability to assess how much they know (Perfect & Stollery, 1993) and thereby increase the confidence-accuracy relationship. If both familiarity and recollection, and hence overall memory quality, were enhanced by the recall phase, we could expect to see a mirror effect in the false alarm and hit rates, with false alarms decreasing and hits increasing at the same time (Reder, Angstadt, Cary, Erikson, & Ayers, 2000). However, a mixed effects model looking at the effect of condition alone on hits (results displayed in Appendix 7A) reveals that no significant increase in hits was produced. Reder et al. (2000) suggest that only the hits portion of the mirror effect is produced by familiarity. Thus, this pattern of overall results suggests that only recollection was enhanced by the recall phase, as intended. Evidence that recollection may have increased despite the fact that there was no differential effect of presentations is provided by a mixed effects analysis of the effect of condition alone on false alarms, which reveals that the recall condition did produce significantly fewer false alarms than the recognition-only condition. This finding suggests that the effects of condition on the confidence-accuracy relationship may have been produced by increased recollection after all.

This raises the question of why the number of presentations did not interact with condition as expected in affecting false alarms and hits. One possible explanation is that recollection was high enough in both conditions to counteract increased familiarity produced through increased presentations. As recent evidence suggests that recollection is continuous, and may be utilised in different ways for different types of decision (Ingram et al., 2012; Mickes et al., 2013), it is possible that recollection was already at a high level in the recognition-only condition, but still increased further following the recall phase. The influence of presentations may not have been sensitive to this increase because recollection was already high enough in the recognition-only condition for participants to realise the increased presentations were the source of heightened familiarity. It is likely that participants would attempt to use recollection as much as possible in this task in order to complete it efficiently, since there were no new words, and therefore participants would realise that familiarity might be misleading as all the words would be words they have seen (Malmberg, 2008; Malmberg & Xu, 2007). Therefore although presentations and condition did not interact to affect false alarms or hits, the finding of decreased false alarms, but not increased hits in the recall condition lends support to the idea that the manipulation may have increased recollection further in this condition.

As the experiment was designed to enhance rather than impair the availability of recollected evidence, these results have more direct application to improving the ability of confidence to predict accuracy. While the previous experiments in this thesis help to elucidate a potential underlying cause of the positive-negative difference, they do not suggest whether knowledge of this cause can lead to methods of reducing the difference, as they only account for methods which may impact on recollection to decrease the confidence-accuracy relationship. Although Experiment 7 did not provide evidence that the degree to which recollection is used is responsible for the changes in the confidence-accuracy relationship, it did provide evidence that the confidence-accuracy relationship for negative decisions can be improved, and the positive-negative difference reduced.

The finding that the positive-negative difference is reduced by the recall phase if replicable would mean that confidence could potentially be used to indicate the accuracy of negative decisions, perhaps not with the same expectations as for positive decisions, but in a similar way as for positive decisions. Prior research in the eyewitness identification field indicates that for some types of recognition decisions, confidence only usefully indicates the accuracy of positive decisions (e.g., Sauer, Brewer, Zweck, & Weber, 2010; Sporer, 1992;

Weber & Varga, 2012). The results of the word recognition experiments in this thesis also suggest that confidence is a weaker predictor of accuracy for negative than positive decisions in a variety of tasks. Hence even where confidence can be used to predict the accuracy of negative decisions it is a much weaker predictor than for positive decisions, and therefore likely to be ignored in applied contexts. Thus if attempting to estimate the likelihood that a given recognition decision is correct, one would have to assume a different relationship between confidence and accuracy depending on decision type, perhaps only taking confidence into account if the decision was positive. Hence the finding that a prior recall phase can reduce the difference in the confidence-accuracy relationship between positive and negative decisions may mean that encouragement of recollection prior to a recognition test would allow a more similar relationship to be assumed when accuracy of a given decision is estimated, at least to the extent that confidence can be usefully taken into account for negative decisions. Manipulation to increase availability of recollection may have potential to increase the confidence-accuracy relationship for negative decisions, thus making it a useful predictor of accuracy, in tasks individuals normally complete relatively automatically using familiarity, such as an eyewitness identification-like task or single face recognition where only one stimulus face has been studied (Malmberg, 2008; Malmberg & Xu, 2007). As this experiment only tested the potential for a recall phase to strengthen the confidence-accuracy relationship in an associative recognition task, it is not yet certain whether this manipulation could play the role of making confidence a meaningful predictor of accuracy in tasks where it may sometimes currently not be. In addition the specific manipulation of a recall phase can only be used in an associative task as the recall test requires the presence of a cue. Hence a logical next step for future research will be to investigate ways of encouraging recollection in item recognition tasks and test the impact of these manipulations on the confidence-accuracy relationship. However for eyewitness identification and associative face recognition tasks where faces are paired with names or occupations, asking participants to bring the studied face to mind prior to completing the recognition test is a relatively simple addition which future studies may attempt.

Before the recall phase manipulation is considered useful for improving assessors' ability to predict accuracy, however, it is important to consider whether the manipulation has any impact on accuracy overall. Prior research has had mixed results concerning the effect of prior testing on subsequent memory performance. Some evidence in the area of recall memory has shown that an earlier test does not improve later recognition (Darley & Murdock, 1971), or that it can decrease recognition accuracy if only certain items have been recalled (Verde, 2004) or produce false recognition if the information has been incorrectly recalled (Roediger & McDermott, 1995). On the other hand some studies have found that being tested (specifically using a recall test) on memory for studied items improves later recognition (Chan & McDermott, 2007; Lockhart, 1975). Therefore the recall test may have an impact on the accuracy of decisions in the subsequent recognition test in the paradigm used for Experiment 7. If the recall phase reduces accuracy on the recognition test, then despite improvement in the confidence-accuracy relationship it would not be useful for applied contexts. Although the mixed effects models shown indicated that the recall phase produced increased accuracy for positive decisions at the lower levels of confidence, this analysis did not test the overall relationship between condition and accuracy. Fixed effects coeffcients for a mixed effects model on the effect of condition on accuracy only are shown in Appendix 7A. The regression coefficient indicates a positive relationship between condition and accuracy showing that the recall condition produced superior accuracy, and the 95% confidence interval indicates that this effect is significant. Thus, the addition of the recall phase significantly affected accuracy in Experiment 7, but did so in a positive direction, providing evidence that the confidence-accuracy relationship for negative associative recognition decisions can be enhanced at no cost to accuracy.

In summary, Experiment 7 provided evidence that it is possible to improve the confidence-accuracy relationship for negative decisions. Future research could build on this finding by testing the recall phase manipulation in tasks which may normally only use limited amounts of recollection, associative face recognition and eyewitness identification. If the recall phase is able to increase the confidence-accuracy relationship for negative decisions in these tasks, this may lead to confidence becoming a useful predictor of accuracy in tasks where it previously has not been. The finding that the confidence-accuracy relationship for negative decisions can be improved is of use for this reason even if increased use of recollection was not the cause, however several alternative explanations are considered unlikely based on the existing literature, and the finding that patterns of hits but not false alarms differed between the conditions overall suggests that increased use of recollection may have been responsible. Thus increasing the availability of recollection for recognition decisions may have potential to strengthen the confidence-accuracy relationship for negative decisions. In addition these findings suggest that asking individuals to first recall the item studied before attempting a recognition decision may be capable of reducing the positive-negative difference. However negative decisions may still never show the same confidence-accuracy relationships as positive decisions, and further research will be needed to define the extent to which confidence-accuracy relationships for both decision types may be made similar.

References

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. Journal of Experimental Psychology-Learning Memory and Cognition, 33(2), 431-437. doi: 10.1037/0278-7393.33.2.431

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. Journal of Experimental Psychology, 91(1), 66-&. doi: 10.1037/h0031836

Huff, M., Davis, S., & Meade, M. (2013). The effects of initial testing on false recall and false recognition in the social contagion of memory paradigm. Memory & Cognition, 41(6), 820-831. doi: 10.3758/s13421-013-0299-4

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. Journal of Experimental Psychology-Learning Memory and Cognition, 38(2), 325-339. doi: 10.1037/a0025483

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. Journal of Memory and Language, 45(1), 82-102. doi: 10.1006/jmla.2000.2761

Lockhart, R. S. (1975). Facilitation of recognition by recall. Journal of Verbal Learning and Verbal Behavior, 14(3), 253-258. doi: 10.1016/s0022-5371(75)80068-5

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. Cognitive Psychology, 57, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. Memory & Cognition, 35(3), 545-556. doi: 10.3758/bf03193293

Mickes, Laura, Seale-Carlisle, Travis M., & Wixted, John T. (2013). Rethinking familiarity: Remember/know judgments in free recall. Journal of Memory and Language, 68(4), 333-349. doi: http://dx.doi.org/10.1016/j.jml.2013.01.001

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults - one deficit or two? Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46(1), 119-135.

Reder, L.M., Angstadt, P., Cary, M., Erikson, M.A. , & Ayers, M.A. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. Learning, memory, and cognition, 26, 294-320. doi: 10.1037/0278-7393.26.2.294

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

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. Law and Human Behavior, 34(4), 337-347. doi: 10.1007/s10979-009-9192-x

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. European Journal of Social Psychology, 22(2), 157-180. doi: 10.1002/ejsp.2420220205

Verde, Michael F. (2004). The retrieval practice effect in associative recognition. Memory & Cognition (pre-2011), 32(8), 1265-1272.

Weber, N., & Varga, M. (2012). Can a modified lineup procedure improve the usefulness of confidence? Journal of Applied Research in Memory and Cognition, 1(3), 152-157. doi: http://dx.doi.org/10.1016/j.jarmac.2012.06.007

Weber, N., Woodard, L., Williamson, P., & Hogan, A. (2014). How do we recognise what we haven't seen? Different bases for positive and negative decisions. Manuscript submitted for publication.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness non-identifications. Psychological Bulletin, 88(3), 776-784. doi: 10.1037/0033-2909.88.3.776

Appendix 7A: Analyses on the Effect of Condition Alone

Table A1

Fixed Effects Coefficients for the Mixed Effects Model Predicting Accuracy from Condition.

Fixed effect	*b*	SE_b	*95% CI*
Intercept	1.48	0.06	[2.84, 3.09]
Condition	0.11	0.09	[0.023, 0.39]

Note. This analysis included test word (*SD* < .001) as a random slope and intercept.

Table A2

Fixed Effects Coefficients for the Mixed Effects Model Predicting False Alarms from Condition.

Fixed effect	*b*	SE_b	*95% CI*
Intercept	-2.04	0.12	[-2.20, -1.86]
Condition	-0.30	0.15	[-0.53, -0.04]

Note. This analysis included test word (*SD* = .31) as a random slope and intercept.

Table A3

Fixed Effects Coefficients for the Mixed Effects Model Predicting Hits from Condition.

Fixed effect	*b*	SE_b	*95% CI*
Intercept	1.09	0.08	[1.99, 2.31]
Condition	0.02	0.11	[-0.14, 0.28]

Note. This analysis included test word (*SD* < .001) as a random slope and intercept.

CHAPTER 8: GENERAL DISCUSSION

This thesis described a series of experiments designed to investigate the impact of the availability of recollected evidence on the confidence-accuracy relationship of negative recognition decisions. Past research has shown that negative recognition decisions frequently have significantly poorer confidence-accuracy relationships than positive recognition decisions. I suggested that this positive-negative difference is due to neglect of recollected evidence in recognition decisions and associated confidence ratings. By comparing circumstances where recollection was available and not available and measuring the confidence-accuracy relationship, the experiments in this thesis provided evidence that the confidence-accuracy relationship was weaker when recollection was impaired and stronger when recollection was made more available. First I examined tasks which normally use very little compared with a large amount of recollection, followed by experiments that altered the amount of recollection available to individuals during their recognition decisions. Results suggested that a) the confidence-accuracy relationship for negative decisions is stronger in tasks where more recollection is used than in tasks where use of recollection is limited; b) the confidence-accuracy relationship for negative decisions is weaker under conditions where recollection is impaired than conditions where recollection is available and c) the confidence-accuracy relationship for negative decisions is strengthened under conditions where the availability of recollection is increased than in a regular associative recognition test. The response latency-accuracy relationship was sometimes found to follow similar patterns to the confidence-accuracy relationship but a) not all experiments allowed consideration of response latency and b) results were not consistent. In Chapter 7 Experiment 7 suggested that when recollection is encouraged, the positive-negative difference may be reduced through greater strengthening of the confidence-accuracy relationship for negative decisions than positive decisions.

Major Findings

Four major questions relevant to the theory were addressed. The first experiments investigated whether the positive-negative difference generalised to item and associative recognition tasks using words and whether there was evidence that these tasks differed in their confidence-accuracy relationships depending on how much recollection was available. The comparison of item recognition and plurality discrimination in Study 1 and Experiments 2a and 2b provided evidence that the positive-negative difference occurs in both item recognition using words and in plurality discrimination. This demonstrates that the positive-negative difference is not a phenomenon specific to recognition tasks involving faces, and that it occurs in at least two different types of word recognition task. In addition, there was evidence of a positive-negative difference occurring under some conditions across all the recognition tasks and stimuli used in the experiments in this thesis. This tells us that the positive-negative difference does occur in other recognition memory paradigms and with other stimuli, not just face recognition and eyewitness identification. The fact that tasks as different from one another as eyewitness identification and plurality discrimination both show a positive-negative difference supports the idea that the cause of the positive-negative difference lies in processes common to all types of episodic recognition tasks, rather than specific features of the tasks in which it was originally noted. This suggests similarities between the underlying processes involved in eyewitness identification and face recognition and those used in common laboratory recognition tasks, which encourages attempts to link all these tasks in recognition theory, as encouraged by Lane and Meissner (2008). In addition, the comparison of item recognition and plurality discrimination provided evidence that the amount of recollection used in a task affected the confidence-accuracy relationship, as the positive-negative difference was not as large in the associative recognition task as in the item recognition task. Associative tasks are thought to use more recollected evidence a) because it is more readily available due to the presence of cues at test and b) because individuals are more likely to perceive use of recollection as efficient for completion of associative tasks where foils share many features of targets and therefore have high familiarity (Malmberg, 2008). Hence if increased recollection promotes a stronger confidence-accuracy relationship for negative decisions, negative decisions could be expected to display stronger confidence-accuracy relationships in an associative recognition task than an item recognition task, and this was shown in the comparison between Study 1 and Experiment 2.

The findings of comparing the item recognition and plurality discrimination task show that the positive-negative difference can occur in laboratory recognition tasks and is not a function of factors specific to tasks involving faces. This supports the argument of theorists such as Lane and Meissner (2008) who suggested that the eyewitness identification task is unlikely to be completely unique in the cognitive processes used. Lane and Meissner argued that because the task of eyewitness identification is likely to share cognitive processes with other kinds of recognition tasks, studies of other kinds of recognition might yield information that applies to eyewitness identification. The finding that the positive-negative difference occurs in word recognition supports this. Lane and Meissner also encouraged application of theories of recognition to the eyewitness context, and the finding that the positive-negative difference generalises to word recognition and is affected by differences in use of recollection between tasks shows that application of general theories of recognition such as dual-process theory may be a useful method of understanding some eyewitness phenomena.

In contrast some authors (e.g., Turtle, Read, Lindsay, & Brimacombe, 2008) have argued that eyewitness identification phenomena should be studied in contexts more similar to the real world and use more real world data, suggesting that findings from the laboratory, even from studies which attempt to replicate eyewitness identification conditions, may not generalise to eyewitness identification in real world situations. However, Turtle et.al. also emphasised that in order to understand phenomena found in eyewitness identification it is necessary to investigate the cognitive and social processes which underlie them. They suggested that many different phenomena seen in the eyewitness identification field may be caused by just a few underlying processes. The results of the experiments in this thesis suggest that the evidence bases (recollection and familiarity) used for recognition decisions may be one of these underlying processes, and that similar phenomena may be seen in other types of recognition tasks when similar underlying processes occur. Hence in the case of the positive-negative difference, conditions which are specific to eyewitness identification may not be involved. Thus these context-specific factors which Turtle et.al. were concerned may be overlooked in laboratory experiments may not be relevant to the positive-negative difference.

Another implication of the finding that the positive-negative difference generalises to word recognition tasks is that negative decisions may have poorer confidence-accuracy relationships than positive decisions across a wide variety of tasks in which the influence of decision type has not yet been investigated. In addition, changes in the degree to which individuals use recollection in these tasks may influence this relationship. Therefore an important task for future research is to investigate other situations in which confidence may be used to indicate the accuracy of important recognition decisions, for example, security officials in airports recognising criminals. An extension of this is that as confidence is involved in metacognitive judgments of other types, such as deciding whether to report or withhold information retrieved on a recall test (e.g., Koriat & Goldsmith, 1996), other types of meta-memory decisions may be impacted by the poor confidence-accuracy relationship for negative decisions.

In addition, in line with Turtle et. al.'s (2008) suggestion that multiple eyewitness identification phenomena may be explained by a single underlying process, the degree to which recollection is used in decisions may also be instrumental to other well-known phenomena in the field. Some evidence for this idea is provided by Marcon, Susa, and Meissner’s (2009) investigation of the role of recollection and familiarity in the own-race bias. Participants in their study were more likely to falsely identify other-race faces following repetition as a strengthening manipulation than own-race faces, suggesting that participants used recollected detail less for other-race faces. Mixed results have been found in recognition of other types of in-group faces, with some support for greater use of recollection, in, for example, own gender faces (Palmer, Brewer, & Horry, 2013). In addition, Meissner, Tredoux, Parker and MacLin (2005) and Gronlund (2005) provided some evidence that the differences between simultaneous and sequential lineups may be related to the degree to which recollection is used in decisions, with more recollection involved in sequential lineup decisions. On the other hand in a review of studies which found a sequential lineup advantage Palmer and Brewer (2012) found that the increase in accuracy was due to a change in response bias rather than an increase in discriminability, and some studies have found that there is no advantage for the sequential lineup procedure in ability to discriminate between lineups where the culprit is present versus absent (e.g., Mickes, Flowe, & Wixted, 2012). If the accuracy advantage of sequential lineups is not due to a change in individuals’ ability to discriminate seen and unseen faces then use of recollection is not a viable explanation. However given the finding in this thesis that the positive-negative difference can generalise to other types of recognition tasks when recollected evidence is neglected, one implication that warrants further investigation is whether other eyewitness phenomena such as differences between sequential and simultaneous lineups or own and other race identification may generalise to other types of recognition as well. Evidence suggests that this may not be the case with sequential lineups as the increase in accuracy for sequential lineups is likely to be due to changes in bias. However the similarities between eyewitness identification tasks and more standard laboratory recognition tasks would be simple to investigate in future research through comparison of simultaneous and sequential multiple choice recognition tests. In addition, a possibility that the evidence on the own-race bias and the findings of the experiments in this thesis raise is that other phenomena in eyewitness identification such as the own-race bias may be reduced or eliminated if recollection was used more in all decisions. Investigating the effects of manipulations to increase recollection on other phenomena regularly observed in the eyewitness identification literature such as the own-race bias is therefore an interesting avenue for further research.

The second major question that this thesis addressed was whether the confidence-accuracy relationship is weakened when recollection is impaired. Having established in Study 1 and Experiments 2a and b that the positive-negative difference occurs in word recognition and varies between tasks with different levels of expected recollection, in experiments 2-5 I investigated the impact of impairing recollection on the confidence-accuracy relationship. This tells us more specifically whether recollection is involved in the confidence-accuracy relationship. A response deadline was used to impair recollection by restricting the time allowed for decisions, or to allow (or possibly even promote) recollection by forcing participants to wait before responding. Results showed that when recollection was impaired by the response deadline, the confidence-accuracy relationship for negative decisions was weaker than when recollection was unimpaired or promoted by a delay before responding. Experiments 3 and 4 demonstrated this result in two different word recognition tasks. This provides more compelling evidence that recollection impacts on the confidence-accuracy relationship, as direct manipulation of a factor known to affect the availability of recollection was used. Recollected evidence is known to become available later than familiarity evidence (Basile & Hampton, 2013; Hintzman, Caulton, & Levitin, 1998; Hintzman & Curran, 1994; McElree, Dolan, & Jacoby, 1999). Therefore, restricting the time allowed for responding can be expected to reduce the availability of recollected evidence. Hence less recollection was expected to be available in the deadline condition. The significantly weaker confidence-accuracy relationship for negative decisions in the deadline condition than in the delay condition therefore supports the idea that recollection is key to the confidence-accuracy relationship for negative decisions. This result also demonstrates that changes in the confidence-accuracy relationship can occur within the same recognition task depending on manipulations which influence the availability of recollected evidence. Hence there is support for the idea that the differences in the confidence-accuracy relationship between Study 1 (item recognition) and Experiments 2a and 2b (plurality discrimination) were not purely due to other differences between the tasks, but are likely to have been due at least in part to differences in availability of recollection as proposed. Again this has implications for any recognition task in which confidence may be used to predict accuracy, as changes in the amount of recollection used may produce changes in the predictive ability of confidence, particularly for negative decisions. However, although the delay condition could be expected to have encouraged recollection by ensuring that adequate time was available for it to occur, the response deadline manipulation did not directly attempt to increase use of recollection. Therefore, a practical application of the finding that changes in recollection are related to changes in the confidence-accuracy relationship was not yet suggested by these findings, as they do not indicate whether use of recollection may be enhanced, and whether this strengthens the confidence-accuracy relationship. Therefore, the role of my final experiment was to answer these questions.

The third major question addressed was whether making recollection more available could produce a stronger confidence-accuracy relationship for negative decisions than is otherwise found. In Experiment 7, I tested whether negative decisions display improved confidence-accuracy relationships when recollection is encouraged. This tells us that when the availability of recollected detail was increased through administration of a recall test before the recognition test, the confidence-accuracy relationship for negative decisions was strengthened. This provides evidence that it is possible to improve the capacity of confidence to predict accuracy for negative decisions by increasing the availability of recollection and that the confidence-accuracy relationship can be affected in both directions by manipulations that alter recollection. This has implications for improving observers' ability to predict an individual's accuracy from their confidence in recognition decisions in general. If the individual is asked to try to bring to mind the item they saw before looking at a test item, this may increase the availability of recollected detail and allow a post-decision confidence rating to be used as an indicator of accuracy regardless of the type of decision made. Alternatively engaging in the process of explicitly recalling information may make that information more salient for subsequent metacognitive processes, encouraging individuals to use the recalled information when making a judgment. Further research is needed to replicate this procedure with different kinds of recognition tasks and in particular with an eyewitness identification paradigm to determine whether the confidence-accuracy relationship for negative decisions can be improved in different contexts and in the original context in which it was found to be problematic.

A final question that the research in this thesis aimed to answer was whether manipulating the amount of recollection available for decisions would lead to changes in the size of the positive-negative difference. This is important to the practical applicability of the finding that recollection affects the confidence-accuracy relationship for negative decisions, because it determines whether positive and negative decisions still have to be considered differently when using confidence to predict accuracy, or if confidence can be used to predict accuracy for both in a similar way. Although the answers to the other questions tell us whether we can change the poor confidence-accuracy relationship for negative decisions, they do not tell us whether we will then be able to use confidence equally well to discriminate between correct and incorrect positive and negative decisions. If confidence significantly predicts accuracy for negative decisions, this allows us to use confidence to diagnose accuracy for both decision types, but we still may not be able to do this equally for positive and negative decisions. This final question concerns whether we can close the gap between the confidence-accuracy relationships for positive and negative decisions such that accuracy can be discriminated using confidence in the same way regardless of decision type. In Experiment 7 where a recall phase was used to increase the availability of recollection this did occur with the confidence-accuracy relationship for negative decisions moving closer to that for positive decisions in the recall condition, where recollection was strengthened. However there was no change in the size of the gap between positive and negative decisions in the other experiments. Instead, in the other experiments the confidence-accuracy relationship was affected for both decision types when the availability of recollection was manipulated. This may reflect the fact that recollection is normally able to be used in positive recognition decisions due to the presence of the target item as a cue, and hence the response deadline restricting recollection also impaired the confidence-accuracy relationship for positive decisions. In contrast when a delay was enforced before responding, recollection may have been available for negative decisions as well as positive, but still less available than for positive decisions due to individuals’ ability to use the test item as a cue for recollection when making positive decisions. In Experiment 7 when participants had to attempt to recollect the studied item before the recognition test, this may have reduced this advantage for positive decisions by making recollected information for negative decisions more equally available during the recognition test, hence reducing the positive-negative difference. However the recall phase manipulation would need to be repeated with different types of tasks, participants and stimuli to determine whether this result is replicable. Although the positive-negative difference was reduced by this manipulation, visual inspection of the data in the form of the graph (Figure 12) shown in Chapter 7 shows that positive decisions did still display a stronger confidence-accuracy relationship even in the recall phase condition. This implies that although it may be possible to increase the confidence-accuracy relationship for negative decisions, and even to make them more comparable with positive decisions, it is still necessary to be aware that the relationship for negative decisions is likely to be weaker when using confidence to indicate accuracy. Some alternative explanations for the results of Experiment 6 are considered in a later section.

Overall the experiments supported the idea that recollection influences the confidence-accuracy relationship for negative decisions. Study 1 and Experiments 2a and 2b demonstrated that the confidence-accuracy relationship is stronger in tasks where more recollection is available and necessary for accurate task completion. All of the experiments demonstrated that a positive-negative difference occurs in the confidence-accuracy relationship in recognition tasks other than those involving faces. Experiments 4 and 5 indicated that restricting individuals' ability to use recollection weakens the confidence-accuracy relationship for both decision types, and Experiment 7 demonstrated that when recollection information is made more available, the confidence-accuracy relationship is strengthened. Taken together these findings suggest that the degree to which recollected evidence is used in decisions has significant impact on the confidence-accuracy relationship for both positive and negative decisions. Positive decisions may normally have a stronger confidence-accuracy relationship than negative decisions due to the fact that even in item recognition tasks, the test item can serve as a cue for recollection of details about studied stimuli, but not unstudied stimuli.

Alternative Explanations for the Findings

Meta-memory Judgments and Overall Accuracy

Manipulations intended to change recollection may also affect overall memory quality, and therefore overall accuracy. Changes in the confidence-accuracy relationship can sometimes be explained by changes in overall accuracy (Kelley & Sahakyan, 2003; Perfect & Stollery, 1993). Thus, it is important to consider that changes in overall accuracy may be difficult to disentangle from changes in recollection in their effect on the confidence-accuracy relationship. For example, Perfect and Stollery (1993) found that the difference between older and younger adults' ability to evaluate their memory reports was due to an overall difference in their recall performance. Kelley and Sahakyan (2003) also suggested that memory monitoring was related to ability to recollect event details. In Kelley and Sahakyan's study dividing attention reduced both the amount of information participants in their study could recollect and the ability of their confidence ratings to distinguish between correct and incorrect information volunteered. In the case of Kelley and Sahkahan's results this could also suggested be due to changes in recollected information, since their result directly relates to recollected details. Table 26 shows the effect of condition on accuracy in each of the experiments in this thesis, along with the respective effects of condition on false alarms and hits. In all experiments except for Experiment 2, overall accuracy was superior in the conditions expected to allow or improve recollection. Hence changes in overall accuracy and changes in recollection are related. However, this may be explained by increases in recollection rather than suggesting that recollection is not responsible for changes in memory monitoring.

There are two reasons why overall accuracy and the confidence-accuracy relationship may be related. The first is statistical, in that changes in overall accuracy may affect outcome

Table 26

Fixed Effects Coefficients for Models Predicting Accuracy from Condition and False Alarms and Hits from Condition in Each of the Experiments Attempting to Manipulate Recollection.

Experiment, Level	*b*	*SE*	*95% CI*
Overall Accuracy			
2, Plurality Discrimination. 1, Delay condition	0.09	0.07	[-0.01, 0.25]
3 Plurality Discrimination, 2, Delay condition	0.68	0.14	[0.41, 0.89]
4 Audio versus Text task, Delay condition	1.02	0.11	[0.88, 1.25]
5, Divided Attention Combined Data, Full attention	0.08	0.13	[-0.08, 0.18]
6, Recall Phase Experiment, recall-first condition	0.11	0.09	[0.03, 0.39]
False Alarms			
2, Plurality Discrimination. 1, Delay condition	-0.11	0.10	[-0.33, 0.03]
3 Plurality Discrimination, 2, Delay condition	-0.63	0.17	[-0.91, -0.44]
4 Audio versus Text task, Delay condition	-0.75	0.17	[-1.02, -0.51]
5, Divided Attention Combined Data, Full attention	-0.03	0.19	[-0.46, 0.52]
6, Recall Phase Experiment, recall-first condition	-0.30	0.15	[-0.58, -0.02]
Hits			
2, Plurality Discrimination. 1, Delay condition	-0.35	0.11	[-0.54, -0.09]
3 Plurality Discrimination, 2, Delay condition	0.39	0.18	[0.08, 0.71]
4 Audio versus Text task, Delay condition	0.44	0.13	[0.14, 0.79]
5, Divided Attention Combined Data, Full attention	0.09	0.13	[-0.14, 0.37]
6, Recall Phase Experiment, recall-first condition	0.02	0.11	[-0.17, 0.25]

measures of the confidence- accuracy relationship. However, mixed-effects analyses effectively control for the overall accuracy of each participant and for differences in overall accuracy between individual stimuli and positive compared with negative stimuli. Thus, there should be no confounding effect of overall accuracy on the effects of recollection on the confidence-accuracy relationship statistically in the experiments in this thesis. The second reason why overall accuracy and memory monitoring may be related is that individuals with superior memory quality are more capable of accurately monitoring their memory. This can be explained from either a single-process or dual-process theory point of view. Single process theories predict this because the familiarity levels of target items and distractor items overlap more for poor, memory than clear memory. This in turn effects individuals' confidence ratings as more clearly separated target and distractor familiarity distributions lead to better discrimination between studied and unstudied stimuli. However, a dual-process account suggests that. as memory quality decreases, the ability to recollect details is reduced. It is this reduction in recollection that causes the poorer monitoring. Kelley and Sahakyan (2003) noted that increased ability to recollect details could be responsible for the increase in memory monitoring in their study. Thus according to dual-process theory, the relationship between overall accuracy and the confidence-accuracy relationship may be explained by greater availability of recollected evidence. The experiments in this thesis support this conclusion, with overall accuracy and the confidence-accuracy relationship consistently stronger in conditions where recollected evidence was made more available.

Explanations for Patterns of Hits and False Alarms

In the experiments in this thesis changes in false alarms and hits with presentations were used to give an indication of whether recollection was altered by the manipulations, and could have therefore been responsible for changes in the confidence-accuracy relationship.

Under conditions that were expected to restrict recollection the expected changes in hits and/or false alarms with presentations of similar stimuli were observed in four of the experiments. Experiments 3, 4, and 7 demonstrated patterns of false alarms and hits that could be considered consistent with changes in use of recollection. However Experiment 2 (confidence data was not analysed due to no evidence of impairment of recollection) and the divided attention experiments in Chapter 6 did not show such effects. The experiments in which patterns of false alarms and hits did not suggest recollection was changing did not show changes in the confidence-accuracy relationship, and thus were consistent with the theory being tested. However ias the evidence for recollection changing in Experiment 6 was mixed, it is worth considering whether there are alternative explanations for the fact that the confidence-accuracy relationship was still strengthened in the condition expected to increase recollection. As explained in the chapter, recollection was not impaired in Experiment 6, and the task was one which normally elicits some use of recollected evidence. Thus participants in Experiment 6 are likely to have already been using recollection prior to the manipulation to increase recollection. This is likely to have allowed them to avoid false alarms to similar foils in both conditions, hence the lack of difference in the change in false alarms with presentations between the conditions. The overall effects of condition on false alarms and hits in Experiment 6 (shown in Table #) show that the increase in overall accuracy in Experiment 6 was due to fewer false alarms occurring overall in the recall condition than the recognition condition, without any accompanying effect on hits. Hence presentations and condition did not interact in their effect on false alarms, but overall the recall condition produced fewer false alarms with no difference in hits. This implies use of recollection to reject similar foils

(Reder et al., 2000) despite the lack of interaction with presentations.[10] Thus in all experiments where changes in confidence and accuracy were observed depending on manipulations intended to alter recollection, there was evidence that recollection indeed changed. In Experiments 3 and 4 hits were higher overall in the delay condition, as well as false alarms being lower, reflecting a general improvement in accuracy. This pattern of increased hits and decreased false alarms demonstrates the mirror effect (Lane & Meissner, 2008). Higham, Perfect and Bruno (2009) found that the mirror effect was best explained by a Type 2 signal detection model which assumes that more conscious recollection is associated with low frequency words. Hence the explanation for this pattern of results offered by Type 2 Signal Detection Theory suggests involvement of recollection, as does the dual-process theory explanation for increases in accuracy.

Finally, there is evidence that recollection can be used to different degrees, or utilised differently for different parts of a decision or task (e.g., Ingram, Mickes, & Wixted, 2012; Mickes, Seale-Carlisle, & Wixted, 2013). Thus the fact that the number of presentations of

[10] Some authors suggest that this pattern of results may reflect reduced, rather than increased, use of recollection (e.g., Joordens & Hockley, 2000). This is the interpretation supported by Type 2 Signal Detection Theory for the false alarm portion of the mirror effect. The mirror effect occurs when one class of studied items (for example, low frequency words compared with high frequency words) which are usually considered more memorable, show both lower false alarm rates and higher hit rates than the other class of items. Joordens and Hockley (2000) suggest that false alarms are lower because the more memorable class of items has less pre-experimental familiarity, and therefore is more easily differentiated based on familiarity into studied or unstudied items. However, as this theory is designed to explain situations in which one of the groups of stimuli is more memorable than the other, this explanation only accounts for the difference in false alarms in Experiment 6 if the items in the recall condition were more memorable than the items in the recognition only condition. Since the same word list was used in both conditions, and words were randomly allocated for each participant to appear in blocks and as targets or foils, this seems unlikely.

stimuli did not interact with the recall-phase manipulation in Experiment 6 may have been due to the fact that adequate recollected information was available in both conditions to offset the increased familiarity of word pairs that had been presented more times. However, the recall phase still made more recollected evidence available increasing a) the number of stimuli at test for which recollected evidence was available, reducing false alarms overall, and b) the amount of recollection available to aid confidence ratings, producing the strengthened confidence-accuracy relationship.

Generalisability to Eyewitness Identification

Since the positive-negative difference was originally noted in eyewitness identification it is of interest to consider whether the findings in this thesis would generalise to the eyewitness identification field. The experiments in this thesis used university students, whose memory may differ from that of the general population, and words, which may differ from faces. Hence it is debatable whether these findings would generalise to eyewitness identification. University student participants are likely to be representative of the general young adult population in most aspects. As most measures of intelligence consider memory in addition to other skills which may be required for university entry, these participants may have somewhat superior memory. However they would presumably not have differing underlying memory processes. As it is the underlying memory processes that are of most concern for the theory I proposed, it is unlikely that vastly different conclusions would be drawn from a different sample of young adults. Further research would be required with wider samples to confirm this. However, I would not expect the effects to necessarily generalise to an older adult or child population as these groups are known to exhibit differences in memory functioning (e.g., Brewer & Day, 2005; Light, Patterson, Chung, & Healy, 2004; Skinner & Fernandes, 2009). An important point for further research is the extent to which recollection affects the confidence-accuracy relationship in different populations and with different stimuli, and whether use of recollection can be increased with benefit to the confidence-accuracy relationship in child and older adult populations. Kelley and Sahakyan (2003) suggest that older adults have poorer memory accuracy and memory monitoring than younger adults due to being less able to recollect the details of events. This could apply to recognition in that older adults may be less able to use recollected evidence in their decision due to being less able to produce this type of evidence. This in turn could affect their ability to appropriately rate confidence in their decision. Future research could investigate whether older adults can be successfully encouraged to use recollection and whether this improves the extent to which their confidence ratings are related to accuracy. In addition young children show deficits in recollection (Ghetti & Angelini, 2008) which may lead to a weak confidence-accuracy relationship. However some studies indicate that children rely on recollection but have poor ability to monitor whether information that comes to mind at retrieval is correct (Czernochowski, Mecklinger, Johansson, & Brinkmann, 2005). Children have been found to display overconfidence when rating their confidence in eyewitness identifications (Brewer & Day, 2005), and confidence has therefore been considered less useful for predicting the accuracy of childrens' identifications. In children this may be due to inaccurate recollection rather than failure to utilise recollection. Hence it is likely that the findings of this thesis will not apply to children. Encouraging recollection in older adults, on the other hand, may alert individuals to decisions which are based on poor memory, and hence improve memory monitoring. Investigation of this possibility is one avenue forfuture research.

While there may be differences in the memory processes investigated in this thesis for different age groups, it is of interest to consider whether the effects found here using plurality discrimination and word-pair recognition tasks would occur in the eyewitness identification context for samples of a similar age. The line-up task does share features with complex laboratory recognition tasks which makes it likely that similar cognitive processes are required. One such feature identified by Malmberg (2008) is that the target must be identified from among highly similar foils. This is a feature of tasks like plurality discrimination where the difference between the target word and a foil may be as minor as the presence of absence of the letter ‘S’. Foils in police lineups are selected specifically for their match to the eyewitness’s description of the culprit, and hence should strongly resemble the offender (Brewer, Weber, & Semmler, 2005; Wells, 1993). Some evidence suggests that manipulations that are thought to reduce recollection, specifically divided attention, can have the same effect in an eyewitness paradigm, and also reduce confidence, suggesting that recollection is normally used to some extent in an eyewitness identification task (Palmer, Brewer, McKinnon, & Weber, 2010). In addition Palmer et al. found that the degree to which participants indicated they could recollect relevant detail related to the studied face was strongly related to confidence. However, the relationship between the type of evidence used in decisions and the confidence-accuracy relationship for eyewitness identification is not yet clear. As discussed in an earlier section, however, other eyewitness phenomena have been shown to be affected by recollection, such as own-race and own-gender effects (Marcon et al., 2009; Palmer et al., 2013) and the sequential lineup advantage (Gronlund, 2005; Meissner et al., 2005), suggesting that recollection is used in eyewitness identification and may be involved in a number of established eyewitness phenomena. Thus it is likely that the findings of the experiments in this thesis will generalise to eyewitness identification to some extent. A further reason that the effects found in this thesis could be expected to generalise to eyewitness identification some degree is that the task of eyewitness identification shares some important features with associative recognition tasks. Given that the culprit must be associated with a certain context, similar to what is required in an associative recognition task where objects must be recognised in association with each other, it is logical that the confidence-accuracy relationship in eyewitness recognition may be affected by recollection in similar ways to tasks such as plurality discrimination and associative recognition. Associative details are required in both types of task in order to make an accurate decision. In associative recognition using word pairs, memory of the presented pair word is needed to determine whether a word pair has been rearranged. In eyewitness identification, memory of the context a face was seen in is necessary to determine whether a face is the one seen in that context, or is familiar due to being seen elsewhere. Hence it makes sense for eyewitness identification to be affected by the underlying memory processes in similar ways to tasks like those used in the experiments in this thesis. However testing this is a task for future research.

If the findings of the experiments in this thesis do generalise to eyewitness identification, there are several implications. One implication is that differences in performance between eyewitness recognition and other associative recognition tasks might arise from differences in the extent to which familiarity and recollection are used as evidence. Face recognition, in which the positive-negative difference has also been found, is usually viewed as an item recognition task, in which recollection is less critical to performance. However, my explanation for the positive-negative difference suggests that, despite the fact that individuals may make accurate decisions based on familiarity, the positive-negative difference will occur in item recognition, due to the comparative lack of use of recollection, particularly for negative decisions. In two of the cited pieces of literature showing an effect of recollection on other established eyewitness identification phenomena, a face recognition paradigm was used (Marcon et al., 2009; Palmer et al., 2013), and effects of recollection observed, suggesting that although recollection may be used to a lesser degree in face recognition, changes in the amount of recollection used in decisions can still occur and can alter the outcomes. These changes in recollection in the case of face recognition without cues may only be possible for positive decisions, given that individuals have no means of accurately recollecting a studied stimulus to compare with a new test face. This may mean that a positive-negative difference will always be present in this type of face recognition task.

While use of simpler laboratory recognition tasks removes additional factors which may influence eyewitness recognition and face recognition, such as uncertain target-absent base rates, variation in the level of similarity between foils and targets (Brewer & Wells, 2006), or simply the fact that people encode and recognise faces differently from other stimuli (e.g., Anaki, Nica, & Moscovitch, 2011) the identified effects of recollection could still be expected to be present in those tasks, if all recognition processes use both recollection and familiarity to varying extents. Increasingly the literature suggests that both processes are used in all types of recognition decisions to varying degrees (e.g., Ingram et al., 2012). Therefore, it can be expected that manipulations which affect these underlying evidence bases will have effects on all types of recognition decisions, just to varying extents.

The important factor in explaining the positive-negative difference in eyewitness recognition is not whether or not recollection is naturally used, but whether or not changes in its use result in changes in the confidence-accuracy relationship. If the eyewitness task is similar to associative recognition tasks, optimal performance in this task may be achieved by approaching it in a similar manner. Dual-process theories usually place different types of recognition decisions on a continuum relating to how likely recollection is to be used; in item recognition, decisions tend to be made based solely on familiarity, whereas in cued-recall, recollection is considered vital (e.g. Malmberg, 2008). It is not yet clear where the eyewitness recognition task falls on this continuum. However as Malmberg suggests, individuals may neglect to use recollection despite the fact that it may improve accuracy or the confidence-accuracy relationship, due to perceiving that it is more efficient to complete the task rapidly using mainly familiarity. Therefore promoting use of recollection in eyewitness identification may be expected to improve the confidence-accuracy relationship for negative decisions by increasing the degree to which negative decisions are based on evidence. Experiments using face recognition and eyewitness identification tasks combined with manipulation of recollection are needed to test this concept. Some ideas for how this may be implemented are covered in the next section.

Practical Implications

The results have several overall practical implications. First, two possible methods of encouraging recollection are raised and each has different implications. The results of Experiment 6 suggest that asking individuals to attempt to bring to mind a studied stimulus before making a recognition decision may have benefits for increasing the capacity to predict accuracy from confidence, especially for negative decisions. Instructing individuals to bring the item they saw to mind is not a difficult strategy to add to most current types of associative recognition tests, since in most cases only the addition of a line of instructions would be required. For example, this could be tested for any laboratory associative recognition task which is structured with study and test phases, simply by adding a screen to computer tasks or a page to paper and pencil tasks before each recognition test, asking participants to bring to mind as clearly as possible the item paired with a certain cue. In eyewitness identification, police conducting lineups may simply be instructed to ask witnesses to bring the culprit’s face to mind as much as possible prior to viewing the lineup. For occupations where recognising criminals in crowds or identifying certain abnormalities on microscope images is critical, individuals might be encouraged to use a strategy of bringing an image of the target to mind prior to observing the setting in which it might appear. Hence if promoting recall of the target prior to a recognition test can improve the ability of confidence to predict accuracy it is relatively practically applicable.

One possible problem with asking individuals to recall an item they saw prior to recognition tasks is that in some fields a testing effect has been found whereby recall of a specific stimulus can impair later memory for similar items (e.g., Dewhurst, Barry, Swannell, Holmes, & Bathurst, 2007). Hence if the situation requires that the individual remember details of a number of similar items, recalling to mind one of those items may impair memory for the others. An example of this in the eyewitness context is unlikely, but possible if a series of individuals are involved in a crime who are related or otherwise of similar appearance, and the witness is required to identify all of these people. Future research would need to investigate the extent to which such a testing effect can occur with faces, so that if this does occur, it may be taken into consideration in the rare instance that a series of individuals of similar appearance are seen by a witness. A testing effect may be more problematic when dealing with own-race biases, as witnesses may view a group of individuals of another race as appearing more similar to one another than they would a group of individuals of their own race (Meissner & Brigham, 2001). Hence it is important for further research to consider not only whether a testing effect might cause a problem for implementing a recall stage before recognition tests with faces in general, but whether there are any differences when dealing with cross-race identifications.

A second possible method of encouraging recollection which the experiments in this thesis have raised is implementing a delay before responding on recognition tests, allowing time for recollected evidence to accumulate. However this may be more difficult to use in an applied context, since it may not always be possible to alter recognition testing methods to allow for a delay before responding. For example, recent evidence shows that speeded confidence ratings can be a more accurate way to obtain positive eyewitness identifications (e.g., Brewer, Weber, Wootton, & Lindsay, 2012) than regular recognition decisions. If the speeded aspect of this procedure is the key feature in improving accuracy, implementing a delay to promote recollection would not be reconcilable with this evidence. However, as the speeded confidence ratings were compared with regular identification decisions, and not with non-speeded confidence ratings, the role of the speed of the ratings was not clear in this study, and allowing individuals more time may not interfere with the accuracy of the procedure. Sauer, Brewer, and Weber (2008) used a similar paradigm but compared non-speeded confidence ratings with control and still found an advantage, suggesting that the speeding may not be the key component of the procedure. Thus allowing more time for recognition decisions may have the capacity to increase recollection, and thus, the confidence-accuracy relationship for negative decisions, without interfering with the procedure's success. The ratings in this alternative procedure avoid the individual making a positive or negative decision. The identification is instead based on the confidence rating given to each face. This type of decision involves confidence that a face was studied, which may vary from confidence that a decision is correct, which is the focus of this thesis. Increasing the diagnostic value of post-decision confidence is still of interest, as in other situations where recognition memory for faces is important, such as identification of a criminal in a real-world situation rather than a lineup, the confidence ratings type of test may not be possible. In these situations post-decision confidence in the accuracy of decisions may still be important in indicating the likelihood that a decision-maker is correct. There are additional situations, however, where the delay manipulation may be difficult to implement as a means of increasing use of recollection. There may be other procedures where speeding is important, the recall phase may be a more useful manipulation to encourage use of recollection in more varied tasks as it may be done before the recognition test, allowing the recognition test to take whatever form is considered best practice at the time. It may also be possible to use a version of the recall phase manipulation which is implemented after individuals have made their recognition decision, but before they rate their confidence in that decision, such that recollected evidence used in the confidence rating is increased. For example, following their recognition decision, individuals could be asked to bring the item they saw to mind and think about how well it matches the item they just responded to. Hence it may be possible to use the manipulations which led to increased recollection in this thesis with other types of recognition paradigms including eyewitness recognition.

One way in which the recall phase manipulation from Experiment 6 might be useful in an eyewitness identification paradigm is to increase the usefulness of ratings for negative decisions in the modified lineup procedure described above. It would be useful for future research to address the issue of whether the recall phase manipulation can usefully combine with eyewitness research on deadlining to increase the availability of recollected evidence for lower confidence ratings which constitute negative decisions (for example, participants can respond with 0%, which is described as certainty the individual is not the culprit). Such confidence ratings might tend to be used by some participants as an indication that they have no evidence in mind to suggest the face is that of the culprit, rather than that they have evidence that the face mismatches the culprit. Individuals can often vary in their understanding of how to realistically use confidence ratings to represent the likelihood that an answer is correct when guessing (Bornstein & Meissner, 2008; Williamson, Weber, & Timmins, 2012). Adding a recall phase may make individuals aware of this distinction and encourage confidence ratings in the middle of the scale, indicating a guess, for situations where they cannot recollect any detail about the appearance of the culprit. For example, participants could be asked to rate the vividness of their recall for the offender prior to seeing the lineup. This would allow observation of the effects of recollected evidence being available to lineup decisions without requiring changes to the procedure.

Implications of Response Latency Results

Latency results were analysed in all experiments to determine whether they followed similar patterns to confidence-accuracy as expected and to provide data for future modelling of the time course of recognition decisions. Results were mixed with latency sometimes showing similar effects to confidence (that is, a stronger response latency-accuracy relationship in the condition expected to include more recollected evidence) but other times not showing a positive-negative difference when confidence did so. Excluding the experiments which manipulated response timing, there was evidence for a positive-negative difference in response latency only in the divided attention experiments overall and in Experiment 6. There was no evidence of a positive-negative difference in response latency in Study 1 or Experiments 2a and b. In both the divided attention experiments overall and in Experiment 7 the response latency-accuracy relationship was stronger in the conditions expected to make more recollection available. However in the case of the divided attention experiments, this was in the absence of an effect of the manipulation on confidence or evidence that the manipulation changed recollection. This tells us that the response latency-accuracy relationship does not necessarily follow the same patterns as the confidence-accuracy relationship and may not respond in the same manner to manipulations intended to increase the strength of the confidence-accuracy relationship. This is potentially due to differences in the optimal time course of positive compared with negative recognition decisions. In earlier chapters the concept was discussed that recollected evidence takes longer to accumulate than familiarity evidence (e.g., Basile & Hampton, 2013; Hintzman et al., 1998; Hintzman & Curran, 1994; McElree et al., 1999). I have also proposed that the most accurate negative decisions require recollection, since contrast between a recollected stimulus and a test stimulus provides evidence for a negative decision, while lack of familiarity only provides lack of evidence for a positive decision. In contrast, highly accurate positive decisions are sometimes possible without recollection since familiarity can provide evidence that an item has been seen, and therefore provides evidence for a positive decision. Hence it is likely that the most rapid negative decisions are less accurate than those made once recollected evidence is available, while the most rapid positive decisions may be more accurate than those for which evidence accumulates more slowly. However for tasks such as plurality discrimination where targets and foils are so similar as to be equally familiar, the response latency-accuracy relationship for positive decisions may also vary. The results of the experiments in this thesis as well as this potential explanation suggest that response latency is unlikely to be able to be used in the same way to predict the accuracy of positive and negative decisions in all tasks. Further research is needed to determine whether response latency is a useful indicator of the accuracy of decisions if used differently for different response types and different tasks. However response latency should not be assumed to predict the accuracy of negative decisions until more is known about the differences between decision types.

General Conclusions

In summary, the experiments reported in this thesis demonstrated three new findings regarding the impact of the availability of recollection on the positive-negative difference in confidence. First, changes in the availability of recollected evidence results in changes in the confidence-accuracy relationship, generally for both decision types. Thus, the theory that recollection influences the confidence-accuracy relationship is supported. Second, the positive-negative difference in confidence occurs in a variety of recognition tasks, and therefore is likely to be due to underlying cognitive processes common to all recognition tasks. Third, the confidence-accuracy relationship for negative decisions can be strengthened by encouraging use of recollected evidence. These findings have a number of implications for theories of recognition and their practical applications. First, since the positive-negative difference does occur in a variety of tasks, decision type is a key factor which should be considered in theories of recognition, as it may change the relationship between other variables often used as indicators of underlying cognitive processes. Decision type is also important for practical applications where confidence may be used to indicate the accuracy of a decision. Confidence is likely to be a stronger indicator of accuracy for decisions where more recollected evidence is used, and for negative decisions, may not be related to accuracy unless at least some recollection is used. In addition a number of new questions which may be pursued in future research have been identified. Since increasing use of recollected evidence can improve the confidence-accuracy relationship for negative decisions, further research may investigate ways of increasing recollection in more applied tasks such as eyewitness lineups and face recognition. In addition, informing witnesses about the role of recollection may assist them in deciding when they should be giving a *don't know* response rather than a negative decision. Key focus points for future research include the investigation of the effect of recollection in recognition paradigms with pictorial or face stimuli and investigation of the relationship between memory quality/vividness of recalled detail and the confidence-accuracy relationship.

References

Anaki, D., Nica, E. I., & Moscovitch, M. (2011). Automatic aspects in face perception evidence from mandatory processing of distractor facial components. Experimental Psychology, 58(1), 4-18. doi: 10.1027/1618-3169/a000061

Basile, B. M., & Hampton, R. R. (2013). Recognition errors suggest fast familiarity and slow recollection in rhesus monkeys. Learning & Memory, 20(8), 431-437. doi: 10.1101/lm.029223.112

Bornstein, Brian H., & Meissner, Christian A. (2008). Basic and applied issues in eyewitness research: A Münsterberg centennial retrospective. Applied Cognitive Psychology, 22(6), 733-736. doi: 10.1002/acp.1478

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. Psychiatry, Psychology and Law, 12(1), 119(110).

Brewer, N., Weber, N., & Semmler, C. (2005). The identification test: Lineup composition. In N. Brewer & K. D. Williams (Eds.), Psychology and Law (pp. 200-202). New York: The Guilford Press.

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. Psychological Science, 23(10), 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G.L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. Journal of experimental psychology. Applied, 12(1), 11-30.

Czernochowski, D., Mecklinger, A., Johansson, M., & Brinkmann, M. (2005). Age-related differences in familiarity and recollection: ERP evidence from a recognition memory study in children and young adults. Cognitive Affective & Behavioral Neuroscience, 5(4), 417-433. doi: 10.3758/cabn.5.4.417

Dewhurst, S. A., Barry, C., Swannell, E., Holmes, S. J., & Bathurst, G. L. (2007). The effect of divided attention on false memory depends on how memory is tested. Memory & Cognition, 35(4), 660-667. doi: 10.3758/bf03193304

Ghetti, Simona, & Angelini, Laura. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. Child Development, 79(2), 339-358. doi: 10.1111/j.1467-8624.2007.01129.x

Gronlund, Scott D. (2005). Sequential lineup advantage: contributions of distinctiveness and recollection. Applied Cognitive Psychology, 19(1), 23-37. doi: 10.1002/acp.1047

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. Journal of Experimental Psychology-Learning Memory and Cognition, 35(1), 57-80. doi: 10.1037/a0013865

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. Memory & Cognition, 26(3), 449-462. doi: 10.3758/bf03201155

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. Journal of Memory and Language, 33(1), 1-18. doi: 10.1006/jmla.1994.1001

Ingram, K. M., Mickes, Laura, & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. Journal of Experimental Psychology-Learning Memory and Cognition, 38(2), 325-339. doi: 10.1037/a0025483

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. Journal of experimental psychology. Learning, memory, and cognition, 26(6), 1534-1555.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. Journal of Memory and Language, 48(4), 704-721. doi: 10.1016/s0749-596x(02)00504-1

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. Psychological Review, 103(3), 490-517. doi: 10.1037/0033-295x.103.3.490

Lane, S. M., & Meissner, C. A. (2008). A 'middle road' approach to bridging the basic-applied divide in eyewitness identification research. Applied Cognitive Psychology, 22(6), 779-787. doi: 10.1002/acp.1482

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. Memory & Cognition, 32(7), 1182-1193. doi: 10.3758/bf03196891

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. Cognitive Psychology, 57, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own- versus other-race faces. Psychonomic Bulletin & Review, 16(1), 99-103. doi: 10.3758/pbr.16.1.99

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. Journal of Experimental Psychology-Learning Memory and Cognition, 25(3), 563-582. doi: 10.1037//0278-7393.25.3.563

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces - A meta-analytic review. Psychology Public Policy and Law, 7(1), 3-35. doi: 10.1037/1076-8971.7.1.3

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. Memory & Cognition, 33(5), 783-792. doi: 10.3758/bf03193074

Mickes, Laura, Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. Journal of Experimental Psychology-Applied, 18(4), 361-376. doi: 10.1037/a0030609

Mickes, Laura, Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/know judgments in free recall. Journal of Memory and Language, 68(4), 333-349. doi: 10.1016/j.jml.2013.01.001

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. Law and Human Behavior, 36(3), 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., & Horry, R. (2013). Understanding gender bias in face recognition: Effects of divided attention at encoding. Acta Psychologica, 142(3), 362-369. doi: 10.1016/j.actpsy.2013.01.009

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. Acta Psychologica, 133(2), 137-145. doi: 10.1016/j.actpsy.2009.11.002

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults - one deficit or two? Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46(1), 119-135.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. Journal of Experimental Psychology-Learning Memory and Cognition, 26(2), 294-320. doi: 10.1037/0278-7393.26.2.294

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. Journal of Experimental Psychology-General, 137(3), 528-547. doi: 10.1037/a0012712

Skinner, E. I., & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults under divided attention. Psychology and Aging, 24(1), 211-216. doi: 10.1037/a0014177

Turtle, J., Read, J. D., Lindsay, D. S., & Brimacombe, C. A. E. (2008). Toward a more informative psychological science of eyewitness evidence. Applied Cognitive Psychology, 22(6), 769-778. doi: 10.1002/acp.1481

Wells, G. L. (1993). What do we know about eyewitness identification? The American psychologist, 48(5), 553-571.

Williamson, P., Weber, N., & Timmins, S. (2012). The role of intuitive statistical knowledge in confidence-accuracy calibration: How people make confidence judgments when guessing. In A. M. Columbus (Ed.), Advances in Psychology Research (Vol. 95, pp. 27-50). Hauppauge NY: Nova Science Publishers.