A GOAL ACTIVATION ACCOUNT OF CONFIDENCE JUDGMENTS

by

Kevin Zish
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Psychology

Committee:

______________________________	Director
______________________________	
______________________________	
______________________________	Department Chairperson
______________________________	Program Director
______________________________	Dean, College of Humanities and Social Sciences
Date: ________________________	Spring Semester 2020 George Mason University Fairfax, VA

A Goal Activation Account of Confidence Judgments

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Kevin Zish
Master of Arts
George Mason University, 2014
Bachelor of Arts
The College of New Jersey, 2011

Director: Tyler Shaw, Associate Professor
Department of Psychology

Spring Semester 2020
George Mason University
Fairfax, VA

DEDICATION

I would like to dedicate this work to my wife Shannon.

ACKNOWLEDGEMENTS

I would like to thank my wife Shannon Zish for her unrelenting support as I completed this research. Shannon and I met when I first moved to Virginia to begin graduate school. As a result, Shannon has experienced the full arc of the endeavor presented here. I cannot express my gratitude enough for helping me think through seemingly impossible problems and providing encouragement as needed, regardless of the volume that the encouragement was supplied. I would also like to thank the many friends and relatives who have made this happen. In particular, I would like to thank the small army of caregivers who watched my daughter Aria so that I could complete this exploration of confidence judgments and human memory. This group of caregivers includes (in no particular order): Deanne McNulty, Ellie McNulty, Denise Roe, Sandy Zish, David Zish, Jennifer Norbury, Mark Norbury, Emily Smith, Natalie Merki, Meredith Munzert, Ileana Alfonso, Brooke Potthast, Assi Spaulding, Sophia Buckman, Delisa O’Brien, Katrina Wunderlich and the general who organized them all—Shannon Zish.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

A GOAL ACTIVATION ACCOUNT OF CONFIDENCE JUDGMENTS
Kevin Zish, M.A.
George Mason University, 2020
Dissertation Director: Dr. Tyler Shaw

Current theories and models of confidence need to be improved to provide a simpler and broader explanation for the relationship between confidence and memory. For example, current theories of confidence posit dozens of features of memory that inform a confidence judgment, making predictions about how confidence changes with novel manipulations difficult. In addition, current models of confidence make robust predictions for tasks with discrete trials but do not specify how a confidence judgment is formed for more complex tasks. In this dissertation I demonstrate that confidence may not be based on a large number of diverse memory features, but rather a more unitary construct, goal activation as defined by Memory for Goals (MFG). MFG is an activation-based model written in the *ACT-R* cognitive architecture. I begin by investigating how confidence changes in procedural tasks. Procedural tasks already have a strong theoretical background rooted in goal activation and provide the opportunity to investigate how confidence changes in a task over time. For Study 1, I demonstrate the existence of confidence carryover which is a novel finding that current models of confidence do not make predictions about. For Study 2, I formalize confidence carryover mathematically with two different procedural tasks. For Study 3, I show that the predicted outcomes of two competing models of confidence judgments can be formed using subtle memory manipulations. The results of Study 3 add a strategic component that is not addressed by any model of confidence judgments. The results of all three studies suggest that goal activation as seen in MFG is the likely source people use when forming confidence judgments.

INTRODUCTION

Recent work in decision-making suggests that when the best choice for a decision is unclear, confidence judgments determine when continued deliberation should stop (Desender et al., 2018, 2019; Dotan et al., 2018). Yet how people make a confidence judgment (i.e. a function of a decision-processes or an evaluation of metamemory) is still an active area of debate making it difficult to predict how confident people will be in different situations. One longstanding consideration for the source of confidence judgments during decision-making is memory (Dunlosky & Metcalfe, 2008) because the relationship between confidence and memory accuracy is reliably positive (i.e. as accuracy goes up so does confidence). The metamemory (Dunlosky & Metcalfe, 2008) and decision-making communities (Pleskac & Busemeyer, 2010) have posited many different features of memory that could be used to make retrospective confidence judgments, including evidence accrual in decision making (drift rate: Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Van Zandt & Maldonado-Molina, 2004), cue utilization and memory accessibility (Gigerenzer et al., 1991; Koriat, 1995, 1997), direct access to memory (Burke et al., 1991; Hart, 1967; King et al., 1980), processing fluency (Kelley & Lindsay, 1993; Koriat, 1997; Van Overschelde, 2008), answer fluency (Jacoby et al., 1989; Kelley & Lindsay, 1993; Whittlesea & Leboe, 2003), frequency (Brase et al., 1998; Gigerenzer et al., 1991), and strength (Clark, 1997; Douglas L. Hintzman, 1986), among others.

Given the many features of memory found to influence confidence, how do people make a confidence judgment?

Many of the features of memory that seem to influence confidence judgments have been explained using *signal detection theory (SDT)*. According to *SDT* a confidence judgment begins with a choice that is determined by a signal sampled along a distribution of evidence from memory (Balakrishnan & Ratcliff, 1996; Ferrell, 1995; Kepecs et al., 2008; Peterson & Birdsall, 1953; Tanner Jr & Swets, 1954; Treisman & Faulkner, 1984; Van Meter & Middleton, 1954; Wallsten & González-Vallejo, 1994). A decision criterion is set along an overlapping distribution of noise and signal plus noise. According to *SDT*, confidence is the scaled distance between the signal and the decision criterion. Stronger, more fluent, or primed memories are farther from the criterion and have higher confidence (cite).

Like *SDT*, I believe that confidence is indicative of the strength of a memory. In this dissertation I investigate whether the memory representation used to form confidence judgments is goal activation, which has been explored extensively in the *ACT-R* cognitive architecture (Anderson, 2007; Anderson et al., 2004). For example, when activation is high confidence is also high (Aguiar et al., 2016; Zish et al., 2015). Through a series of studies, I show that activation can provide a robust explanation of the behavior of confidence judgments in memory tasks because activation captures many of the features described before (i.e. activation increases for stronger, more fluent, or primed memories). Importantly I am not trying to take well-defined theories of confidence like *SDT* and repackage them as activation. Instead, using activation to explain confidence judgments has a major benefit, namely that activation has a very particular form, an equation. *ACT-R* suggests that the activation of a memory element is defined by the relationship shown in **Equation 1** between the base-level activation B_i and it's history of use, where t is the time since the j^{th} use in competition with decay d (Equation 2.2 in Anderson, Bothell, Lebiere, & Matessa, 1998).

$$B_i = ln(\sum_{j=1}^{n} t_j^{-d})$$

Equation 1. Base-level activation equation.

Equation 2 shows a simplified activation equation of a memory element m that is defined by how many times a memory is rehearsed n over time T (in addition to priming from other elements and error).

$$m = ln(n/\sqrt{T})$$

Equation 2. Simplified activation equation.

Equation 2 makes the relationship between rehearsal and time explicit. Goals that have been rehearsed many times in the recent past have more activation than goals that have been rehearsed fewer times or rehearsed in the distant past.

If confidence judgments do arise from activation, then activation-based models can provide clear predictions for how confidence judgements would change in memory tasks. The *Memory for Goals* (*MFG*: Altmann & Trafton, 2002) model is an activation-based model originally described in *ACT-R* that describes unique patterns of performance in goal-based behavior. According to *MFG,* activation of a memory element is determined by four components namely the strengthening constraint, the priming constraint, the interference level, and decay.

The strengthening constraint suggests that activation is based on the frequency and recency of retrievals. For example, memories that were retrieved recently or have been retrieved repeatedly will have higher activation than memories that are older or retrieved less often. The priming constraint suggests that elements in memory are connected by associative links. Memories receive activation when other memories they are linked to are retrieved. Altmann & Trafton (2007) showed that when steps in a task are completed concurrently, cumulative priming builds in subsequent steps through these associative links allowing for cognitive control of the next correct step of a task to be maintained. The interference level is determined by activation of competing elements in memory. For the correct memory to be retrieved, an element in memory must have activation greater than its competitors which define the interference level. Finally, goals also undergo decay. Decay is an important part of forgetting and is indexed by time. The more time that has passed since a goal has received activation (from being retrieved or from associative priming) the lower the activation for the goal.

If confidence is based on activation, it follows that confidence should behave in several systematic ways according to *MFG*. Manipulations that impact activation should also impact confidence in a similar magnitude and direction. The present set of studies was designed to test if confidence varies in the same systematic way as activation. Study 1 uses a procedural task to demonstrate confidence carryover from trial to trial. Confidence carryover can be explained by similar strengthening and priming mechanisms in *MFG* that explain the rate of sequence errors. In Study 2, we demonstrate that confidence judgments can be influenced by earlier trials via a priming mechanism that utilizes associative strength between elements in memory. We formalize the relationship between a priming mechanism and confidence by developing a mathematical model that uses associative strength as a parameter. By fitting the mathematical model to empirical data, we find that following an interruption, confidence can recover over time which the model explains as priming. The same mechanism has been used to explain the recovery of response time following an interruption and uses some of the same formalisms as *MFG*. For Study 3, we use activation to predict the results of a recognition paradigm.

STUDY 1.

Abstract

Current models of memory demonstrate a consistently negative relationship between confidence and error. However, many models are built using experimental paradigms that utilize discrete trials where stimuli are independent of each other. I used a procedural task to measure confidence as participants made different types of sequence errors to investigate confidence carryover effects. I demonstrate that confidence can depend on previous trials in a systematic way that can be explained by an activation-based model.

Introduction

Abundant empirical evidence suggests that confidence judgments have a strong memory component. For example, when memory is manipulated (i.e. interruptions, perceptual judgment tasks, targets on categorized lists, etc.), confidence consistently has a negative relationship with error such that higher error rates result in lower confidence for retrospective memories (Baranski & Petrusic, 1998; DeSoto & Roediger, 2014; Dunlosky & Metcalfe, 2008; Pleskac & Busemeyer, 2010; Roediger & DeSoto, 2014; Zish, Hassanzadeh, McCurry, & Trafton, 2015). However, many models and experiments measuring confidence consider the contribution of memory only from the item just retrieved (DeSoto & Roediger, 2014; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Van Zandt & Maldonado-Molina, 2004; Vickers, 2014). These models are silent or do not explicitly predict if or how confidence changes (i.e.

confidence carryover) for memories that are linked, such is the case with procedural tasks which include making coffee (Botvinick & Bylsma, 2005; Botvinick & Plaut, 2004; Cooper & Shallice, 2006; Cooper & Shallice, 2000), programming a device (Li et al., 2008), making a photo copy (Byrne & Davis, 2006), or filling up a car with fuel (Chung & Byrne, 2008).

For example, to make a cup of coffee, a set of ordered tasks must be completed. First, a cup is produced, then coffee is poured into a cup, followed by sugar and milk. Finally, the coffee, milk, and sugar are stirred. Completing a task out of order is known as a sequence error. Sequence errors are the result of a cognitive system failing to maintain control over the next appropriate action in a task (Baars, 1992; Norman, 1981; Reason, 1990). In general, sequence errors occur more frequent the shorter the temporal distance to the next correct step of the task (Altmann & Trafton, 2007; Altmann, Trafton, & Hambrick, 2014; Botvinick & Bylsma, 2005; Botvinick & Plaut, 2004; Cooper & Shallice, 2006; Cooper & Shallice, 2000; Hiatt & Trafton, 2015; Trafton, Altmann, & Ratwani, 2011). As a result, most sequence errors consist of skipping ahead one step or repeating the step just completed. These two types of errors are more commonly known as perseverations and anticipations (Parasuraman & Davies, 1984; Reason, 1990).

A variety of models suggest a priming component accounts for the pattern of carryover effects seen with sequence errors (Botvinick & Bylsma, 2005; Botvinick & Plaut, 2004; Cooper & Shallice, 2006; Cooper & Shallice, 2000). One such model is the *Memory for Goals* model (*MFG*: Altmann & Trafton, 2002). *MFG* is an activation-based model built in the *ACT-R* cognitive architecture (Anderson, 2007; Anderson et al., 2004).

According to *MFG*, activation of a memory element is determined by three components, namely the strengthening constraint, the priming constraint, and the interference level.

The strengthening constraint suggests that activation is based on the frequency and recency of retrievals. For example, memories that were retrieved recently or have been retrieved repeatedly will have higher activation than memories that are older or retrieved less often (Anderson, 2007; Anderson et al., 2004). The priming constraint suggests that elements in memory are connected by associative links (Hodgetts & Jones, 2006). Memories receive activation when other linked memories are retrieved (Cades et al., 2011). Altmann & Trafton (2007) showed that when steps in a task are completed concurrently, cumulative priming builds in subsequent steps through these associative links allowing for cognitive control of the next correct step of a task to be maintained. Finally, the interference level is determined by activation of competing elements in memory (Thomson et al., 2017; Trafton et al., 2011). For the correct memory to be retrieved, an element in memory must have activation greater than its competitors and it is the activation of the most active competitor that defines the interference level.

Using an interruption paradigm, Trafton et al. (2011) modeled the frequency of anticipations and perseverations in a procedural task using *ACT-R*. According to their model of sequence errors, perseverations errors are well explained by the strengthening constraint and decay. In the context of a procedural task, items in memory retrieved more recently have higher activation than items retrieved in the more distant past. As items in memory decay, activation for items in memory retrieved temporally close to each other become increasingly similar. Following an interruption, decay can accumulate for a just completed step of a task such that activation for the just completed step and its competitors is indistinguishable.

Anticipation errors are the result of decay interacting with the priming constraint (Altmann et al., 2014). Interruptions effectively cut off activation that was otherwise passed to subsequent parts of the task through associative links. Upon returning to the task, decay of elements in memory makes activation for the next correct step in the task similar to its competitors.

The *MFG* account for sequence errors from Trafton et al. (2011) provides a basis for how carryover effects influence performance from one part of a procedural task to another. The well-established negative relationship between error and confidence (Ariely et al., 2000; Baranski & Petrusic, 1998; Dougherty, 2001; Garrett, 1922; Johnson, 1939; Nelson, 1990; Vickers, 2014) suggests that explanations for changes in the accuracy of memory should also apply to confidence. If confidence is a function of memory, then *MFG* makes a series of predictions.

As shown in Trafton et al. (2011), *MFG* predicts a decrease in sequence errors the farther a step is from the next correct step of a task. According to the strengthening constraint, activation is lower for items in memory that are from the distant past and provide less competition with the correct goal of the task. Therefore, I should expect confidence will be lower for more temporally distant steps in the procedure that are repeated. In addition, activation is lower for items in memory from the distant future and confidence should be lower for steps in the procedure that are prematurely anticipated. The priming constraint suggests that these items in memory have lower activation due to cumulative priming through associative links. Similar to Trafton et al. (2011), I use an interruption paradigm in a procedural task to elicit sequence errors to investigate the hypothesis that confidence decreases for more temporally distant sequence errors from the next correct step of the task.

Methods and Materials

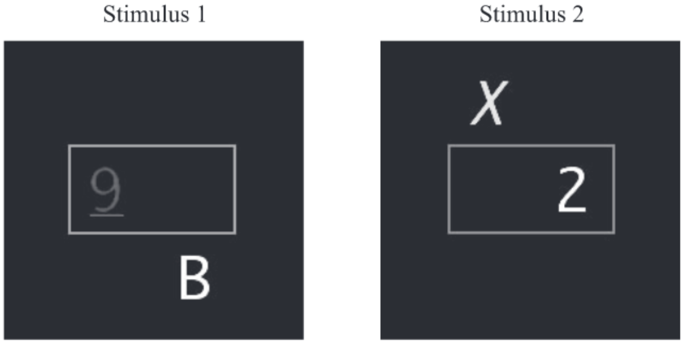
Participants

One hundred and fifty-five George Mason University undergraduates participated for partial course credit.

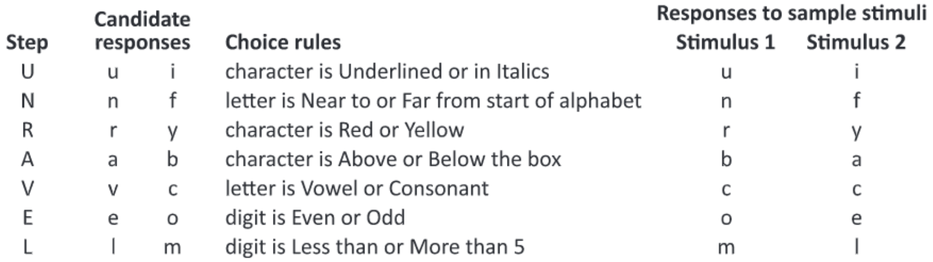
Primary Task

The UNRAVEL task was adapted from Altmann, Trafton, & Hambrick (2014). The UNRAVEL task has seven rules each represented by a letter (e.g. U, N, R, A, V, E, L). Participants are presented with one number and one letter at the same time. Each letter and number have certain characteristics that change from trial to trial such as color, font, position, etc. Participants are instructed to keep the UNRAVEL rule in memory, interpret what characteristic of the stimuli they are asked to identify, analyze the stimuli, and using the keyboard to submit what characteristic they identified (**Figure 1**).

(a) Sample stimuli for UNRAVEL task:



(b) Choice rules and candidate responses for UNRAVEL task, and responses to stimuli in (a):



Step	Candidate responses		Choice rules	Responses to sample stimuli Stimulus 1	Stimulus 2
U	u	i	character is Underlined or in Italics	u	i
N	n	f	letter is Near to or Far from start of alphabet	n	f
R	r	y	character is Red or Yellow	r	y
A	a	b	character is Above or Below the box	b	a
V	v	c	letter is Vowel or Consonant	c	c
E	e	o	digit is Even or Odd	o	e
L	l	m	digit is Less than or More than 5	m	l

Figure 1. Example of the UNRAVEL task from Altmann et al. (2014).

For example, the U action in UNRAVEL prompts participants to identify if a number or letter is underlined or italicized. If a letter or number is underlined, they press the “u” key on the board. If the letter is italicized, they are instructed to press the “i”. After they submit their response participants will be presented with a new stimulus. They will search the stimulus for a characteristic prompted by the N action. The N action prompts participants to determine if the letter is near (“n”) or far (“f”) from the beginning of the alphabet. Participants continue to proceed through the UNRAVEL rules. Once completed, participants wrap around to the U action. The goal is to complete the rules in order and correctly identify the prompted characteristic for each stimulus.

Each action in UNRAVEL has a different set of keys associated with a response. As a result, the keystrokes reveal what action participants think they are on.

Interruption Task

Participants were interrupted 11.85% of the time using a process described in Altmann et al., 2014. After an UNRAVEL response was submitted, the UNRAVEL task was occluded and participants were asked to type in a series of 14 letters into a box. Once the letters were typed in correctly the UNRAVEL task was revealed again. Participants were asked to resume the UNRAVEL task where they left off.

Confidence Question

Participants received a confidence question after completing an UNRAVEL action following half of the interruptions and an equal number of the control trials. The task was replaced with a question that asked: “How confident were you that you just chose the correct step during the UNRAVEL task? Enter your choice on a scale from 1 to 6, with 1 being *least confide*nt and 6 being *most confident*.” I used an even numbered scale so that participants could not provide an ambiguous middle response although recent work suggests any number of scales (i.e. 1-4, 1-20, 1-100) and would be equally valid (Tekin & Roediger, 2017). The participant typed in their response into a text field. After submitting their response, the participant was returned to the UNRAVEL task

Procedure

Participants filled out an approved IRB consent form. The task was first described using screenshots.

Participants were given a practice session where each rule of UNRAVEL was explained. They were exposed to all elements of the task including interruptions and confidence questions. Participants were shown that they could hit a key to access a list of the UNRAVEL rules at any time.

Measures

I calculated the frequency and mean confidence of each type of correct/error response.

Results

One hundred and fifty-five participants completed 42,442 UNRAVEL actions. I treated each action as a trial. There were 4,925 confidence judgments. I analyzed the data from trials associated with a confidence judgement for the 114 participants who correctly identified the next step of the task 95% of the time. In addition, recall that my hypothesis is that confidence should change systematically across sequence errors. Because sequence errors increase after an interruption, I focused my analysis on the 1,852 trials immediately following an interruption.

Frequency Distribution

Error behavior was coded relative to the correct step of the trial. UNRAVEL is a procedural task that must be completed in a specific order and that the response type by the participant allows the experimenter to determine what rule the participant thinks is correct. Therefore, if the next correct step of the task was the A step, but participants completed the trial with rules from the R step, the action was coded as a -1 perseveration. More egregious perseverations, such as selecting the N or U step were -2 or -3 perseverations, respectively. If a participant selected the V step, the action was coded as a +1 anticipation. Completing the E or L step was coded as a +2 or +3 anticipation, respectively. An example of responses and the associated error type with that answer when the correct step is A can be seen in **Table 1**.

Table 1. Example of responses and error types in the UNRAVEL task.

Response	U	N	R	A	V	E	L
Steps Away from the Correct Answer	-3	-2	-1	0	1	2	3

The distribution of responses can be seen in **Figure 2** and is similar to the pattern seen in Trafton et al. (2011).

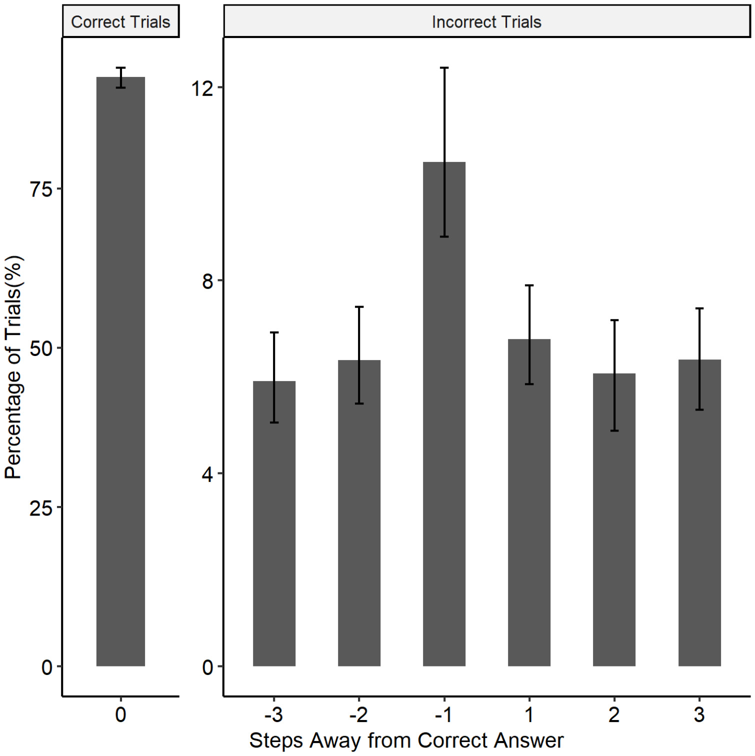


Figure 2. Distribution of responses. Error bars are 95% confidence intervals.

Confidence

Consistent with previous research, an unbalanced ANOVA with type III sums of squares shows that confidence was significantly higher for correct trials (M = 5.22, SD = .83) than incorrect trials (M = 3.93, SD = 1.60), F(1,180) = 50.84, MSE = 70.81, $p < .05$, η_p^2 = .22. An unbalanced ANOVA was used because some of the 114 participants did not make any errors following an interruption.

If confidence is the evaluation of activation, then *MFG* predicts that confidence will vary by error type. *MFG* predicts that errors psychologically further from the correct step (e.g. -3 and +3) should have less activation than errors closer to the correct step (e.g. -2 and

2 or -1 and +1) due to an interaction between primacy, recency, strengthening, and decay. Specifically, items in memory that have been retrieved in the more distant past should have lower activation than items retrieved recently. Therefore, lower activation for answer type -3 perseverations should result in lower confidence that gradually increases with a peak at the next correct step of the task and then decreases again until answer type +3 anticipations. The quadratic pattern described was revealed by a polynomial contrast for mean confidence judgments across answer types and can be seen in **Figure 3**, $F(6,212) = 11.77$, $MSE = 18.79$, $p < .05$; $\eta_p^2 = .25$; Quadratic: $t = -4.24$, $p < .05$. An unbalanced ANOVA was used to evaluate the overall polynomial contrast because not all participants made an error at each sequence error step. The significant quadratic pattern suggests confidence varies systematically with lower confidence for sequence errors farther from the correct step of the task.

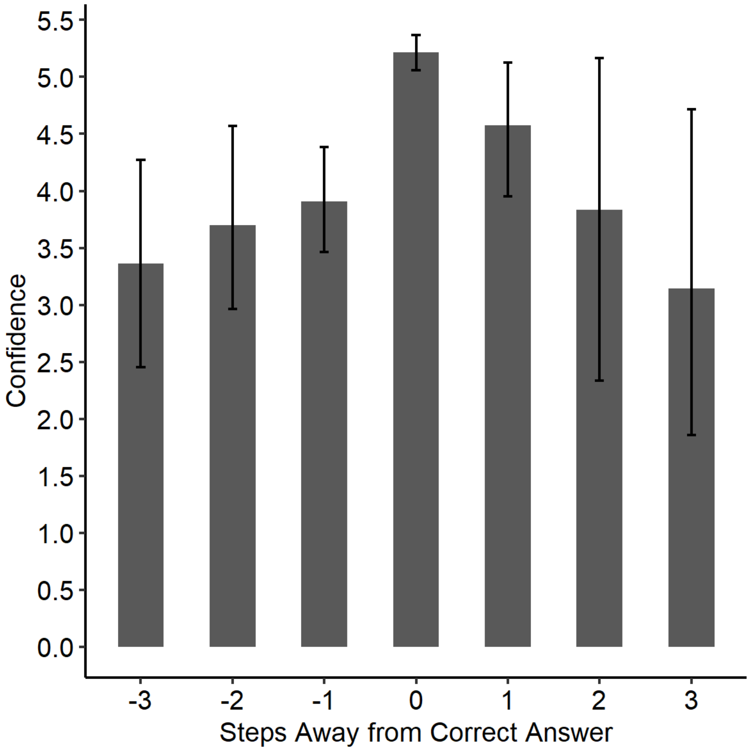


Figure 3 Confidence rating by answer type for all steps. Error bars are 95% confidence intervals.

As **Figure 3** suggests, there is a distinct pattern of confidence for 0, -1, -2, and -3 steps away from the correct answer. Specifically, confidence is higher for correct trials and then decreases with perseveration distance; a linear contrast within an unbalanced ANOVA with type III sums of squares supports this perception, $F(3,182) = 21.94$, $MSE = 31.23$, $p < .05$, $\eta_p^2 = .27$; Linear: $t = 4.87$, $p < .05$. A similar pattern emerges for anticipation errors that are 0, 1, 2, and 3 away from the correct answer: confidence is higher for correct steps and then decreases with anticipation distance, $F(3,143) = 12.47$, $MSE = 13.74$, $p < .05$, Linear: $t = -5.28$, $p < .05$.

Discussion

This study used a procedural task to produce sequence errors so that confidence carryover effects could be investigated. Consistent with other studies, our results show that confidence discriminates between correct and incorrect trials (Baranski & Petrusic, 1998; Pleskac & Busemeyer, 2010; Vickers, 2014). Unique to this study, our results provide evidence of confidence carryover, where the process used to form a confidence judgment in one trial is linked to the process used to form a confidence judgment in a subsequent trial. Specifically, two polynomial contrasts show that confidence judgments decrease systematically for both perseverations and anticipations according to the distance the error was from the next correct step.

MFG readily provides a process that explains why confidence decreased for more temporally distant errors from the next correct step of the task. According to the strengthening and priming constraints, activation for the just completed step (-1) and the about to be completed step (+1) are the highest and that higher activation is why the largest proportion of errors congregated around the correct step of the task both here and in Trafton et al. (2011). The strengthening constraint explains that activation decays for goals retrieved in memory in the more distant pass. Similarly, the priming constraint explains that associative links are formed for items retrieved in close proximity. Associative links provide cumulative priming activation to future goals, providing less and less priming activation to goals farther in the future than nearer goals. Based on this study, confidence systematically changes based on primacy, recency, and decay as described by MFG. Therefore, confidence should systematically decrease for goals farther in the future because psychologically more distant goals should have lower activation

Clearly, confidence judgments are based on some feature of memory. However, the exact feature of memory has led to much debate in the literature. Current models of confidence, such as the *2 Stage Dynamic Signal Detection* (*2DSD*: Pleskac & Busemeyer, 2010) are remarkably robust and provide an excellent explanation for a number of findings in the confidence literature such as discrimination between correct and incorrect trials. However, *2DSD*s interpretation of our findings with confidence carryover is less clear. *2DSD* currently uses a diffusion process (Ratcliff, 1978) to explain decision making and suggests that the quality and speed of evidence accrual (drift rate) inform confidence judgments. Specifically, higher drift rates result in higher confidence. However, *2DSD* does not have a clear explanation for the curvilinear decrease in confidence found here and would require an additional priming component allowing drift rate from one decision to carry over to another decision. The results of this study show the beginnings of a simpler explanation: confidence judgments are based on activation such that higher activation leads to higher confidence.

STUDY 2.

Abstract

I investigated the time course of confidence after a task interruption, focusing on participants' level of confidence that they were resuming the task correctly. A new mathematical model of confidence was built which shows why confidence decreases on trials immediately following an interruption and then gradually increases. The model uses associative strength as a parameter and helps to formalize the relationship between activation and confidence.

Introduction

Many models of confidence only consider the contribution of the item in memory just retrieved and suggest decision-making is a unitary process (Baranski & Petrusic, 1998; DeSoto & Roediger, 2014; Kiani, Corthell, & Shadlen, 2014; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Vickers, 2014). This independent view of confidence judgements is likely an artifact of the types of tasks commonly used to study confidence such as perceptual stimuli (Baranski & Petrusic, 1998; Pleskac & Buseymeyer, 2010; Yu et al. 2015), list-learning (Desoto & Roediger, 2014), recognition memory (Ratcliff & Starns, 2009), or general knowledge questions (Pleskac & Buseymeyer, 2010). However, many real-world tasks are procedural, which rely on the associative strength between elements in memory to determine the next correct step of the task (Hiatt & Trafton, 2015; Trafton et al., 2011). At best, current models of confidence are silent about a priming mechanism that would influence confidence judgments via associative strength. In this study I investigate how confidence changes in procedural tasks. I turn to model-based activation experiments that have been able to demonstrate how priming in procedural tasks influences other well-understood measures, namely, response time. The example I use is from Altmann & Trafton (2007), who used an interruption paradigm to investigate response time while resuming a procedural task. Using a similar approach, I interrupted participants during a procedural task and measured confidence during the resumption process. By measuring confidence during the resumption process, I can determine if well-studied priming mechanisms that drive procedural task behavior can also account for confidence judgments.

Altmann & Trafton (2007) had participants complete a dynamic decision-making task, periodically interrupting participants after they completed an action. The time to resume the task after the interruption (i.e. resumption lag: Trafton, Altmann, Brock, & Mintz, 2003) was measured up to ten actions after the interruption was complete. The authors found that interruptions increased resumption time after the interruption. However, participants did not immediately recover after the first action following an interruption. Instead, the data showed that resumption lag followed a curvilinear pattern of recovery wherein the response time for each action after the interruption was faster than the preceding action.

Altmann & Trafton (2007) developed a mathematical model to explain how memory and attention recovers following an interruption. The model suggests that actions in procedural tasks are associatively linked. These associative links result in a feed-forward priming mechanism that plays a critical role in facilitating the selection of the next action in a task. When an action is completed, priming from the completed action is added to all following actions. The total priming of an element at position *p* is represented by:

$$A(p) = -1 + \sum_{i=1}^{p} assoc^{i-1}, assoc < 1$$

Equation 3. Equation for activation at position *p*, (Altmann & Trafton, 2007).

where in **Equation 3** *assoc* is the amount of associative strength between elements of the task measured between 0 and 1. An *assoc* value of 0 assumes no associative strength between elements in memory and collapses the equation into the value of the *F* parameter only. Positive values of *assoc* suggests that elements in memory are associated and is used to calculate the amount of activation that has accumulated from previous elements.

In practice, the action directly after the current action receives the most priming. Subsequent actions receive lower and lower amounts of priming. When participants complete an uninterrupted task, an element receives small amounts of priming from each associatively linked action before it. Interruptions effectively cut off that priming making it so that priming for the action to be resumed is lower compared to if it had not been interrupted.

Their primed-retrieval model for choice RT is represented by **Equation 4** where *F* is a scaling parameter representing non-decisional processes multiplied by $e^{-A(p)}$:

$$RT(p) = F * exp[1 - \sum_{i=1}^{p} assoc^{i-1}]$$

Equation 4. Time course of recovery for RT, (Anderson, 2007; Anderson et al., 2004).

The work in Altmann & Trafton (2007) is built on the activation-based *Memory for Goals* (*MFG*: Altmann & Trafton, 2002) model which uses *ACT-R* (Anderson, 2007; Anderson et al., 2004). *ACT-R* is a cognitive architecture that formalizes many of the findings in cognitive psychology related to human behavior. I used the activation-based properties of *MFG* to make predictions about the relationship between choice RT and confidence. For example, goals that have been rehearsed many times in the recent past have more activation than goals that have been rehearsed fewer times or rehearsed in the distant past.

Goals also undergo decay. Decay is an important part of forgetting and is indexed by time. The more time that has passed since a goal has received activation (from being retrieved or from associative priming) the lower the activation for the goal. Interruptions decrease activation by decreasing the probability that a goal can be retrieved, increasing the amount of time between rehearsal, or some combination of the two.

In the study that follows I measured confidence after a participant made a choice about what they believe was the next correct step of the task. If confidence is influenced by a priming mechanism, it follows that confidence should behave in several systematic ways according to activation dynamics following from *MFG*. First, confidence is a strong function of memory and should decrease after an interruption because interruptions decrease activation. This finding has already been demonstrated by Aguiar, Zish, McCurry, & Trafton (2016) and Zish, Hassanzadeh, McCurry, & Trafton (2015).

Second, if confidence judgments are associatively linked, confidence should increase in a curvilinear pattern following an interruption. Confidence should decrease after an interruption because interruptions effectively eliminate cumulative priming from previous trials. Following an interruption, associatively linked elements in memory provide a cumulative priming mechanism for future trials. Cumulative priming from previous elements in memory that should result in a gradual increase in confidence for later actions as the decision process recovers.

Third, a mathematical model of the time course of recovery for confidence *C* can be built that will match empirical data. The model I propose is shown in **Equation 5**:

$$C(p) = S * exp[-1 + \sum_{i=1}^{p} assoc^{i-1}]$$

Equation 5. Time course of recovery for confidence.

This approach allows me to instantiate my first two predictions, namely that confidence is a strong function of memory and that confidence will increase in a systematic manner after an interruption. *S* in **Equation 5** is different than *F* in **Equation 4** because response time and confidence are on different scales. The parameter *S* is a scaling parameter that represents non-decisional processes. In addition, I changed the cumulative priming equation from *–A(p)* for RT to *A(p)* given that RT decreases after an interruption and confidence increases. I investigate these systematic effects using a procedural task.

Methods and Materials

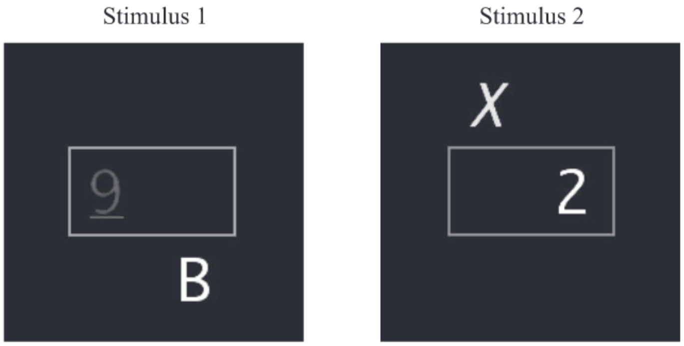
Participants

One hundred and fifty-five George Mason University undergraduates participated for partial course credit.

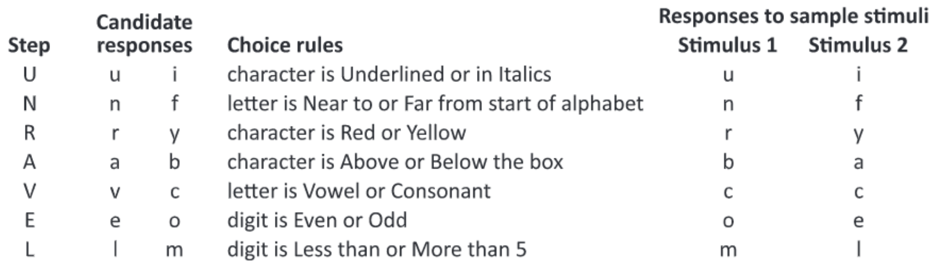
Primary Task

The UNRAVEL task was adapted from Altmann, Trafton, & Hambrick (2014). The UNRAVEL task has seven rules each represented by a letter (e.g. U, N, R, A, V, E, L). Participants are presented with one number and one letter at the same time. Each letter and number have certain characteristics that change from trial to trial such as color, font, position, etc. Participants are instructed to keep the UNRAVEL rule in memory, interpret what characteristic of the stimuli they are asked to identify, analyze the stimuli, and using the keyboard to submit what characteristic they identified (**Figure 4**).

(a) Sample stimuli for UNRAVEL task:



(b) Choice rules and candidate responses for UNRAVEL task, and responses to stimuli in (a):



Step	Candidate responses		Choice rules	Responses to sample stimuli Stimulus 1	Stimulus 2
U	u	i	character is Underlined or in Italics	u	i
N	n	f	letter is Near to or Far from start of alphabet	n	f
R	r	y	character is Red or Yellow	r	y
A	a	b	character is Above or Below the box	b	a
V	v	c	letter is Vowel or Consonant	c	c
E	e	o	digit is Even or Odd	o	e
L	l	m	digit is Less than or More than 5	m	l

Figure 4. Example of the UNRAVEL task from Altmann et al. (2014).

For example, the U action in UNRAVEL prompts participants to identify if a number or letter is underlined or italicized. If a letter or number is underlined, they press the “u” key on the board. If the letter is italicized, they are instructed to press the “i”. After they submit their response participants will be presented with a new stimulus. They will search the stimulus for a characteristic prompted by the N action. The N action prompts participants to determine if the letter is near (“n”) or far (“f”) from the beginning of the alphabet. Participants continue to proceed through the UNRAVEL rules. Once completed, participants wrap around to the U action. The goal is to complete the rules in order and correctly identify the prompted characteristic for each stimulus.

Each action in UNRAVEL has a different set of keys associated with a response. As a result, the keystrokes reveal what action participants think they are on.

Interruption Task

Participants were interrupted 11.85% of the time in a process detailed in Altmann et al. (2014). After an UNRAVEL response was submitted, the UNRAVEL task was occluded and participants were asked to type in a series of 14 letters into a box. Once the letters were typed in correctly the UNRAVEL task was revealed again. Participants were asked to resume the UNRAVEL task where they left off.

Confidence Question

Participants received a confidence question after completing an UNRAVEL action following half of the interruptions and an equal number of the control trials. The task was replaced with a question that asked: "How confident were you that you just chose the correct step during the UNRAVEL task? Enter your choice on a scale from 1 to 6, with 1 being *least confide*nt and 6 being *most confident*." I used an even numbered scale so that participants could not provide an ambiguous middle response. The participant typed in their response into a text field. After submitting their response, the participant was returned to the UNRAVEL task

Procedure

Participants filled out an approved IRB consent form. Participants were given a practice session where each rule of UNRAVEL was explained. They were exposed to all elements of the task including interruptions and confidence questions. Participants were shown that they could hit a key to access a list of the UNRAVEL rules at any time.

Measures

I calculated the mean response time and confidence rating for the first seven trials after an interruption.

Results

One hundred and fifty-five participants completed 42,442 UNRAVEL actions. I treated each action as a trial. There were 4,925 confidence judgments. I analyzed the data from trials associated with a confidence judgement for the 114 participants who correctly identified the next step of the task 95% of the time. In addition, recall that my hypothesis is that confidence is influenced by a priming mechanism following an interruption. Any investigation of a priming mechanism should consider correct trials only since the proportion of correct to error trials increases as people recover from an interruption. Error trials decrease confidence judgments which could be a confounding variable. Therefore, I focused my analysis on the 2,701 correct responses that followed within seven trials after an interruption.

Modeling the Time Course of Recovery for Decisions RT

Not every participant had a confidence question after each of the seven steps after an interruption. I used an unbalanced ANOVA with type III sums of squares to look for differences in RT across trial after an interruption. There was a significant effect of trial after an interruption, $F(6,624) = 8.36$, $MSE = 20.97$, $p < .05$, $\eta_p^2 = .07$. To investigate differences between steps, I compared Step 1 with Steps 2-7 using a repeated measures ANOVA. Trial 1 ($M = 4.21$, $SD = 1.28$) after an interruption was significantly higher than Trials 2-7 ($M = 3.08$, $SD = .96$), $F(1,113) = 92.03$, $MSE = 72.53$, $p < .05$, $\eta_p^2 = .45$. This result replicates the immediate disruptive effects of interruptions (Gillie & Broadbent, 1989; Trafton et al., 2003; Trafton & Monk, 2007).

Replicating Altmann & Trafton (2007) the response time for the primary judgment has a curvilinear pattern of recovery after an interruption. A polynomial contrast revealed the data had both a significant and quadratic shape, Linear: $t = -5.37$, $p < .05$; Quadratic: $t = 2.97$, $p < .05$. Following Altmann & Trafton (2007), I fit the model to the data by estimating *F* and *assoc* for each participant. I varied both the *F* parameter from 0 to 14 (i.e. the range of mean response times in seconds for each trial since the interruption) by .1 and the *assoc* parameter from 0 to 1 by .1 and calculated model fits for each trial after an interruption with every parameter combination. An *RMSE* was calculated for each parameter combination for each participant. Only the lowest *RMSE* for each *F* and *assoc* parameter was used for each participant. The *F* and *assoc* were then averaged across participants to give us a mean *F* of 4.27 (*SD* = 1.22) and a mean *assoc* of .24 (*SD* = .17) that ranged from .1 to .7. All model points are within the 95% confidence intervals of the empirical data. The mean *RMSE* was 1.02 (*SD* = .68) and R^2 was .39 (*SD* = .29). **Figure 5** shows the empirical data for choice RT for each UNRAVEL step after an interruption and predicted choice RT from our model. This replicates Altmann & Trafton (2007).

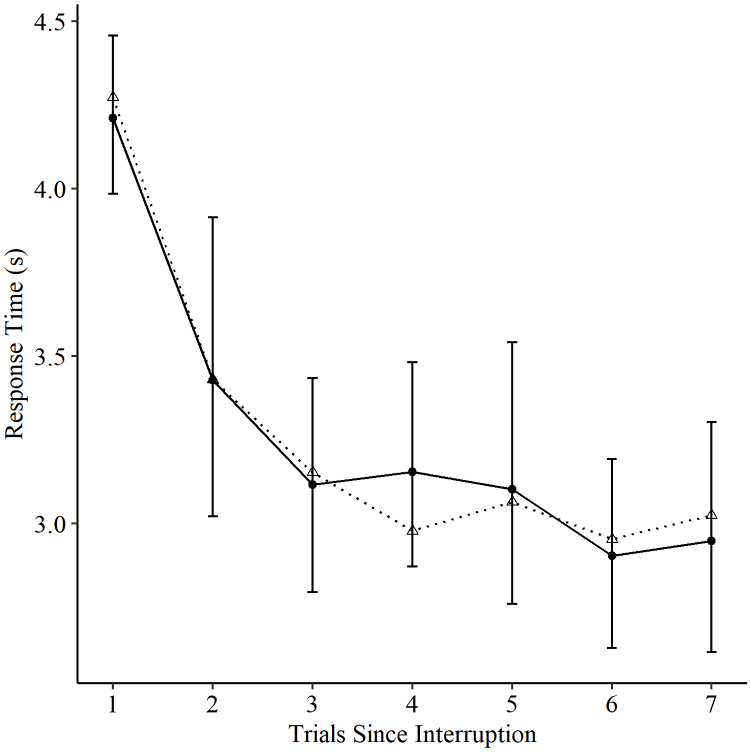


Figure 5. Data (solid circles) and model (dashed triangles) for time course recovery of choice RT. Error bars are 95% confidence intervals data.

To test goodness of fit, runs tests were applied (Bradley, 1968). The runs test examines the signs of the deviations from the model minus the data. If the data or the model were different shapes, the deviations between the data and the model would be systematic. For example, if a set of data showed an increase in some metric and a model showed a decrease in the same metric, then the signs would be different for the beginning of the data, be the same in the middle when the data and model crossed, and then different again. The runs test applied here showed that deviations in the data and the model were the same suggesting they are equal, *t*(154) = .22, *p* = .83.

Modeling the Time Course of Recovery for Confidence

The process to model confidence was identical to choice RT. I varied the *S* parameter from 1 to 6 (i.e. the possible confidence response choices) by .1 and the *assoc* parameter from 0 to 1 by .1 and chose the combination of parameters with the lowest *RMSE*. Surprisingly, participants fell into one of two *assoc* parameter groups. Consistent with current models of confidence 68 participants had a best-fitting *assoc* parameter value of 0 which implies no associative strength between trials and that confidence was determined independently. For the remaining 48 participants the best-fitting *assoc* parameter value was positive and ranged from .1 to .4. A positive *assoc* parameter implies that for 42% of participants, confidence was informed by the associative strength between trials which is not predicted by previous models. I analyze these two groups separately.

To investigate confidence when the *assoc* parameter was a 0, an unbalanced ANOVA with type III sums of squares was used. There was no difference in confidence by trial after an interruption $F(6,362) < 1$ and no pattern of recovery following an interruption; Linear: $t = -1.26$, $p > .05$; Quadratic: $t = .21$, $p > .05$.

The *S* and *assoc* parameters were estimated for each participant and the parameters with the lowest RMSE was retained as the best-fitting model for that participant. Across participants, the mean *S* parameter was 5.49 (*SD* = .78). The mean *RMSE* was .32 (*SD* = .33). All model points are within the 95% confidence intervals of the empirical data. When *assoc* is 0, the model is a straight line with the y-intercept at *S* which makes R^2 undefined. **Figure 6** shows the empirical data for confidence for each UNRAVEL action after an interruption and predicted confidence from our model.

The goodness of fit for the runs test for the model and data were the same, $t(67)= .52$, $p = .60$.

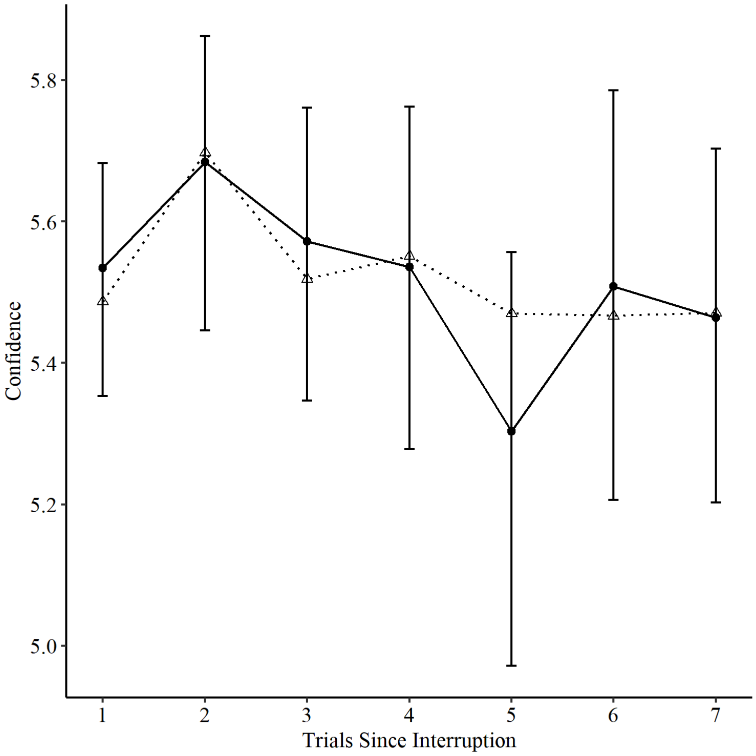


Figure 6. Data (solid circles) and model (dashed triangles) for time course recovery of confidence when the associative strength between trials is 0. Error bars are 95% confidence intervals data.

To investigate confidence when the *assoc* parameter was greater than 0, an unbalanced ANOVA with type III sums of squares showed a difference in confidence by trial after an interruption $F(6,255) = 4.79$, $MSE = 3.75$, $p < .05$, $\eta_p^2 = .10$. To investigate the differences between trials I compared Trial 1 with Trials 2-7. There was a significant difference between Trial 1 ($M = 4.75$, $SD = .79$) and Trials 2-7 ($M = 5.45$, $SD = .76$),

$F(1,45) = 221.50$, $MSE = 11.30$, $p < .05$, $\eta_p^2 = .83$. A polynomial contrast revealed a nearly curvilinear pattern of recovery where confidence decreased after an interruption and increased linearly; Linear: $t = 4.33$, $p < .05$; Quadratic: $t = -1.77$, $p = .08$.

The *S* and *assoc* parameters were estimated for each participant and the parameters with the lowest RMSE was retained as the best-fitting model for that participant. Across participants the mean *S* parameter was 4.68 (*SD* = .88) and the *assoc* parameter was .14 (*SD* = .07). The mean *RMSE* was .33 (*SD* = .31) and mean R^2 was .58 (*SD* = .37). **Figure 7** shows the empirical data for confidence for each UNRAVEL action after an interruption and predicted confidence from our model.

The goodness of fit for the runs test for the model and data showed no significant difference, $t(45)= -1.78$, $p = .08$.

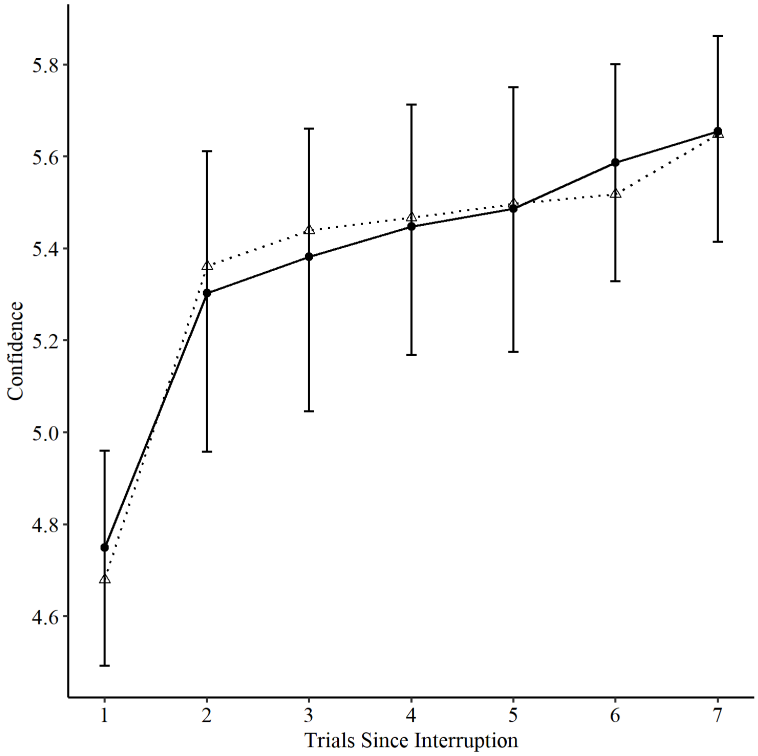


Figure 7. Data (solid circles) and model (dashed triangles) for time course recovery of confidence when associative strength is greater than 0. Error bars are 95% confidence intervals for data.

Discussion

On a procedural task I built two models of complex cognitive processes: decision-making and confidence judgments. I instantiated decision-making using the model from Altmann & Trafton (2007) to model choice RT. Similar to the decision-making model I built a model of confidence that also used an associative strength parameter.

First, I was able to replicate the empirical curvilinear pattern from Altmann & Trafton (2007) and show that RT recovers over time after an interruption. In addition, I fit **Equation 4** of choice RT from Altmann & Trafton (2007) to a new task.

Second, I presented empirical data that shows for 42% of participants, confidence recovers over time. This is a novel finding given that many models and experiments measuring confidence only consider the contribution of the item in memory just retrieved (Baranski & Petrusic, 1998; DeSoto & Roediger, 2014; Kiani, Corthell, & Shadlen, 2014; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Vickers, 2014).

Third, I built a novel model for confidence judgments that fits the data and describes the recovery process of confidence judgments following an interruption. The model uses a cumulative priming mechanism to drive the recovery of confidence. The cumulative priming mechanism suggests that confidence judgments are linked by associative strength and at least in some cases are not independent decisions.

Unique to this study, I did not rely on the common list-learning or perceptual stimuli that have come to dominate the field. Instead I used a complex procedural task which allowed me to demonstrate that in some cases confidence recovers over time. The data shows that for some participants confidence increased gradually after an interruption. My model of confidence explains why confidence increases: some participants made use of a priming mechanism when forming a confidence judgment. This priming mechanism is driven by associative strength between memory elements. As participants resumed the task, priming accumulated in a systematic way which increased confidence judgments.

In contrast, other models assume that confidence judgments are independent of one another (Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Van Zandt & Maldonado-Molina, 2004). The model presented here does not disregard previous independent accounts of confidence judgments. In fact, our model suggests that for 58% of participants there was no associative strength between trials. An *assoc* parameter of 0 implies participants used the most recently retrieved item to form a confidence judgment which was independent from the previous trial. However, our model adds something new: a way to account for a priming mechanism in confidence judgments. A positive *assoc* parameter implies that a large minority of participants utilized associative strength between trials to inform their confidence judgment. In combination with our empirical data, this study provides strong evidence that forming a confidence judgment is not always an exclusively independent process of prior decisions.

Current sequential sampling models assume that confidence judgments are independent of one another (Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Van Zandt & Maldonado-Molina, 2004). However, for 42% of our participants, confidence was influenced by the associative strength between trials. None of the sequential sampling models discussed here specify a priming mechanism that would allow confidence to be informed by associative strength between trials. Alternatively, *MFG* has a readily available mechanism for the effect of interruptions on confidence judgments, goal activation.

STUDY 3.

Abstract

In this study we compare two competing models of confidence judgments in a 2AFC task. The *relative account* of confidence assumes that people use the difference in strength between a target and a lure to form a confidence judgment. Alternatively, the *absolute account* of confidence assumes that only the strength of the target is used to form a confidence judgment and the contribution of the lure is ignored. Using a recognition memory paradigm, I manipulate features of memory to encourage participants to use one account of confidence over the other. This study introduces a strategic component to confidence judgments based on the context of the stimuli that should be included in future models of confidence judgments.

Introduction

The decision-making and metacognition literature have a long history of using theoretical models to determine how people form confidence judgments. In short, models can be classified into one of two groups based on what evidence is used to produce a confidence judgment. For ease of explanation, we call the first group of models the *relative account* and the second group of models the *absolute account*, although these models have been given other names throughout the literature (winner-take-all: Miyoshi et al., 2018; response-congruent evidence rule: Maniscalco et al., 2016; Miyoshi & Lau, 2020; balance-of-evidence: Vickers, 2014). According to the *relative account*, in a decision-making task the strength of both the choice and its competitors is used inform confidence judgments. Confidence is the difference in strength between a single choice and the unchosen alternatives. In contrast, the *absolute account* suggests only the strength of the choice determines confidence. In this study, I investigate the use of absolute and relative evidence using a recognition memory paradigm to determine if people can produce the predicted effects for both accounts. A demonstration of the effects from the *relative* and *absolute accounts* would suggest a strategy component missing from all known models of confidence judgments.

The *relative account* is supported by *sequential sampling models* (Kiani et al., 2014; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013) such as *2DSD* (Pleskac & Busemeyer, 2010; Yu et al., 2015) and the *Poisson model* (Merkle & Van Zandt, 2006; Van Zandt & Maldonado-Molina, 2004) that show confidence is determined by the *balance of evidence* (Vickers, 2014). In 2-alternative forced-choice (2AFC) tasks, higher confidence is the result of a larger difference in the strength of evidence between a target and a lure (although the same mechanism can apply to *n*-forced choice tasks, see Horry & Brewer, 2016). Support for the *relative account* comes largely from studies involving perceptual stimuli which show a positive relationship between confidence and discriminability (Baranski & Petrusic, 1998; Garrett, 1922; Johnson, 1939).

The *absolute account* is rooted in *SDT* (Balakrishnan & Ratcliff, 1996; Ferrell, 1995; Kepecs et al., 2008; Peterson & Birdsall, 1953; Tanner Jr & Swets, 1954; Treisman & Faulkner, 1984; Van Meter & Middleton, 1954; Wallsten & González-Vallejo, 1994). According to the *absolute account*, judgments are produced by comparing a signal sampled from a distribution of signal plus noise to a decision criterion. *SDT* implies that people make confidence judgments using one element from memory, the choice. If the element in memory sampled from signal plus noise is far from the decision criterion, confidence is high. Previous work using perceptual stimuli showed that confidence changed only with the strength of the choice (Maniscalco et al., 2016; Samaha et al., 2017; Zylberberg et al., 2012). For example, Zylberberg, Barttfeld, & Sigman (2012) manipulated the perceptual differences between random dot motions (Experiment 1) and luminance (Experiment 2) in a 2AFC task. Confidence increased as the strength of the target increased, completely ignoring the contribution of the unchosen alternative.

More recently, there have been several attempts to contrast both accounts using 2AFC recognition memory paradigms. Showing that the same account of confidence occurs for both perceptual and recognition memory tasks could offer one major advantage to either the *relative* or *absolute accounts* of confidence: evidence of a domain-independent mechanism. In one study by Zawadzka, Higham, & Hanczakowski (2017), the authors used the plurals paradigm (Hintzman & Curran, 1994). Participants studied the singular or plural word of a noun (e.g. apple, bananas, etc.). During a study phase, participants were simultaneously presented with a target and a lure. A lure would be the same word studied previously but in the opposite form seen in the study (e.g. apples, banana). At test, participants saw pairs of targets and lures that varied by the number of rehearsals at study. The results of their study supported the *absolute account* of confidence. Confidence in correct trials increased as strength increased, even if the strength of the lure had also increased. One other recent study also found support for the *absolute account* of confidence using scene memorability (Miyoshi et al., 2018). In Miyoshi, Kuwaharac, & Kawaguchia (2018), participants were presented with scenes with low, medium, and high memorability taken from a memorability database (Bylinskii et al., 2015). Participants were presented each scene, one at a time, for 50ms. During a test phase, two images were presented on the screen that varied by memorability. Their results showed that confidence was highest when the strength of the target was highest.

Both studies seemingly support the *absolute account* of confidence for 2AFC recognition memories. Yet, many 2AFC perceptual paradigms support the *relative account* of confidence. How can the two accounts be reconciled? —especially when both the *relative* and *absolute accounts* assume that people can only use one method to make a confidence judgment. We suggest that both models of confidence are correct and the method of determining a confidence judgment is strategic or task driven.

To investigate a possible strategic component, a recognition memory paradigm was used. A recognition memory approach offers the advantage of having access to robust theories of memory such as *ACT-R* (Anderson, 2007) which is a cognitive architecture of findings in psychology that can explain a wide array of cognitive processes. *ACT-R* has a robust explanation of memory retrieval based on activation. The base-activation of any item in memory is shown in **Equation 6** where t is the time since the j^{th} use in competition with decay d (Equation 2.2 in Anderson, Bothell, Lebiere, & Matessa, 1998).

$$B_i = ln(\sum_{j=1}^{n} t_j^{-d})$$

Equation 6. Base-level activation equation.

Equation 7 shows a simplified activation equation of a memory element *m* that is defined by how many times a memory is rehearsed *n* over time *T* (in addition to priming from other elements and error).

$$m = ln(n/\sqrt{T})$$

Equation 7. Simplified activation equation.

Equation 7 makes the relationship between rehearsal and time explicit. Memories that have been rehearsed many times in the recent past have more activation than goals that have been rehearsed fewer times or rehearsed in the distant past.

Like strength, items with higher activation are more likely to be retrieved quickly and accurately from memory than items with lower activation. Calculating activation affords the ability to more precisely predict the strength of an item in memory and the difference in strength between items in a 2AFC task. Calculating activation for each stimulus in a 2AFC task has a major advantage over other less precise methods of varying strength. Other methods can only make assumptions about the absolute and relative strength of items in memory. In contrast, I can use the simplified activation equation to predict the absolute and relative strength of evidence between items when planning conditions. This process was helpful when considering how high or how different activation will be between items and whether people are likely to use relative or absolute strength to inform their confidence judgments.

If people can strategically use the *relative* or *absolute account* of confidence, then it should be possible to manipulate situations where people form confidence judgments in line with either account. Many 2AFC studies show a *relative account* of confidence across a variety of tasks that also demonstrate a consistent relationship with performance metrics such as accuracy (see Pleskac & Buseymeyer, 2010 for a review). For example, when accuracy is high, confidence is also high. When discriminability is low, confidence is also low (Baranski & Petrusic, 1998). However, a few more recent 2AFC experiments show a dissociation between confidence and accuracy. When discriminability is low, confidence is high (Maniscalco et al., 2016; Zylberberg et al., 2012). An inverse relationship between confidence and task difficulty is counter intuitive. If the purpose of confidence judgments is to determine the probability that an answer is correct (Desender et al., 2018, 2019; Dotan et al., 2018) why would participants use a suboptimal strategy to make a confidence judgment?

Some experiments have explored what a dissociation between accuracy and confidence means for typical models of confidence (Desender et al., 2017; Peters et al., 2017; Rahnev & Denison, 2018). While I do not discuss them here, I interpret the existence of a dissociation between accuracy and confidence to mean that the *absolute account* of confidence is not the default strategy for producing a confidence judgment in 2AFC tasks. Instead, participants will use the *relative account* of confidence when the difference in strength between items is large.

In the study that follows, I present participants with a list of nonwords. Nonwords come with an undefined amount of activation in *ACT-R* because it is impossible to remember something that has never been in memory before. Novel stimuli reduce some noise likely seen in other experiments that use previously known items (e.g. nouns or images). I varied the frequency of presentation, giving some groups of nonwords higher activation than other groups. During a test phase, participants were asked to determine which of two nonwords was more familiar. According to the *relative account*, confidence in the target of a 2AFC task should decrease as the difference in strength between the target and the lure decreases. The opposite is true for the *absolute account*. Confidence should increase in the target as strength increases.

In Study 3a, I attempt to elicit the *relative account* of confidence by making the difference in strength between nonwords large. If participants are using the *relative account* of confidence, then confidence should decrease as the difference in presentation frequency for each nonword decreases. For Study 3b I attempt to elicit the *absolute account* of confidence by reducing the difference in strength between nonwords and increasing the strength of the target from trial-to-trial.

STUDY 3A.

Methods and Materials

Participants

Fifty-one people participated through Amazon's Mechanical Turk.

Design

I used nonwords to reduce the chance that any stimuli would have higher activation than other stimuli prior to the study. A word generator (http://www.neuro.mcw.edu/mcword/) created a total of 64 nonsense words in the form of constrained tri-grams. A constrained trigram is defined as "a specific three letter combination (trigram) in a specific position, in a specific length of a word ("MCWord: An Orthographic Wordform Database"). That is, the 'sta' in 'stage' is considered the same 'sta' as in 'staff' but is different from the trigram in 'stay'." Constrained trigrams provide words that are more pronounceable in the English language than unconstrained words because the combinations of three letters occur in the same location as actual words in English. All nonsense words were 6 characters long (e.g. "absure", "poldon", "laddly", "memain").

For this study, a 1-factor repeated measures design with four levels was used where participants were tested on nonwords that had been previously rehearsed 0 times (two new words), 0 and 5 times (one new word and one well-rehearsed word), 1 and 5 times, and 4 and 5 times (two well-rehearsed words). Notice that while the maximum number of rehearsals is the same for 0_5, 1_5, and 4_5, the relative number of rehearsals is different. In this study, an *absolute account* would lead to confidence that was the same across all conditions, while a *relative account* would lead to a decrease in confidence as the practice with lures increased.

Procedure

Participants were presented with 48 nonsense words during a study phase either 1, 4, or 5 times. When a nonsense word was presented on a computer monitor, participants typed in the nonsense word into a text box one at a time. To ensure that participants saw and encoded the nonword at least once, the next trial would only present itself if the spelling of the typed word matched the nonsense word.

During the test phase, all the nonsense words from the study phase were presented as part of 32 pairs of nonsense words. Participants were given the instruction to select the word that was the most familiar. One of the nonsense words was presented on the left side of the screen and the other on the right side of the screen. Pairs of nonsense words varied in how often each nonsense word was presented during the study phase.

The 32 pairs of nonsense words were split into four equal groups based on the frequency of their presentation during the study phase. Each nonsense word in a pair was represented in the data by a number which exemplified how often each word was seen during the study phase. For example, the first group contained eight pairs of nonsense words that were new and had not been seen during the study phase. The purpose of this group was to encourage full use of the confidence scale by having one condition with maximum uncertainty. This pair of two new words during the test phase was characterized by two zeroes (i.e. 0_0). The other three groups contained eight pairs of nonsense words studied zero and five times (Pairing: 0_5), studied one and five times (Pairing: 1_5), and studied four and five times (Pairing: 4_5). During the test phase, each group of nonsense words was block randomized such that participants saw only one example from each group before seeing another example of a pairing.

We manipulated activation by changing the presentation frequency of each item in a pair of nonsense words. Specific to Study 3a, we attempted to elicit the *relative* account of confidence by keeping the presentation frequency of the target trials the same throughout the study phase (i.e. the 5 in 0_5, 1_5, and 4_5). According to the simplified activation equation, another way to manipulate activation would be to account for the time between an item was presented and when it was tested. In future work we address this simplification by controlling for time which would result in tighter control of activation at test.

Once a participant selected which nonsense word was most familiar, they received a confidence question that asked to rate their confidence on a scale of 1 to 6 with 1 being “Least Confident” and 6 being “Most Confident”.

Measures

We calculated the mean accuracy for each group of nonsense words as well as the confidence rating for correct trials only. We restricted the analysis of confidence to correct trials only to exclude the possibility that differences in confidence was the result of varying proportions of error trials for each group.

Results

Fifty participants completed the study. We did not include the 0_0 condition in our analyses because there was no correct answer. There were 1,224 trials for analysis.

Accuracy

There was a significant difference in accuracy between the 0_5 (M = 93.38, SD = .13), 1_5 (M = 85.78, SD = .16), and 4_5 conditions (M = 59.07, SD = .17), F(2,100) = 4.60, MSE = 1.16, $p < .05$, $\eta_p^2 = .08$. To determine the shape and direction of trends in the data, we used a polynomial contrast. The polynomial contrast revealed that as the difference in activation between the target and the lure decreased (i.e. 0, 1, 4), the ability to discriminate which word was more familiar decreased linearly, Linear: $t = -11.13$, $p < .05$ (**Figure 8**).

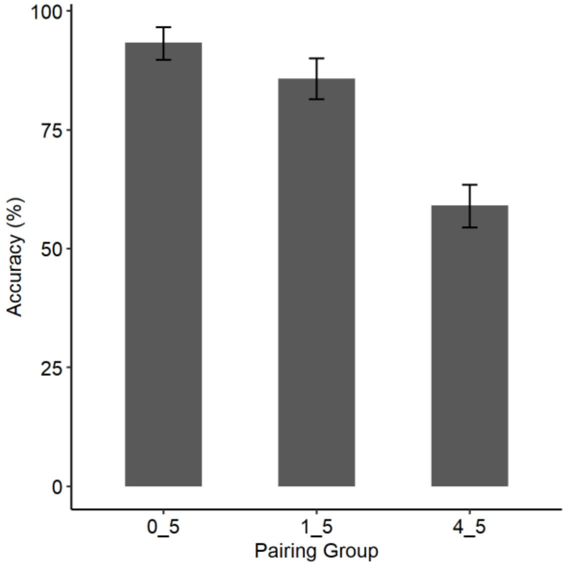


Figure 8 Accuracy for Study 3a. Error bars are 95% confidence intervals.

Confidence

There was a significant difference in confidence for correct trials between the 0_5 (*M* = 5.25, *SD* = .90), 1_5 (*M* = 5.14, *SD* = .85), and 4_5 conditions (*M* = 4.82, *SD* = 1.16), *F*(2,100) = 8.24, *MSE* = 2.49, p < .05, η_p^2 = .14. As revealed by a polynomial contrast and seen in **Figure 9**, when the difference in practice between each pair of words decreased, confidence linearly decreased, Linear: *t* = -2.19, *p* < .05).

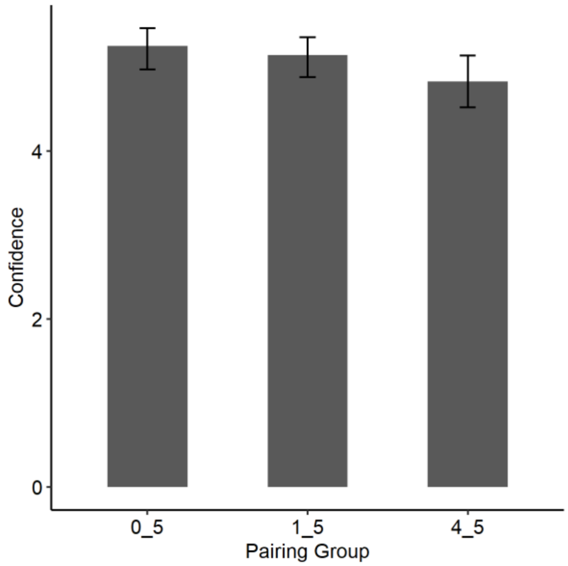


Figure 9. Confidence for correct trials only for Study 3a. Error bars are 95% confidence intervals.

Discussion

In Study 3a, the accuracy data suggests that discriminability of the target decreased when the presentation frequency of the target remained constant, but the presentation of the lure increased. Consistent with the *relative account* of confidence, I found that confidence decreased as discriminability decreased. These results suggest that participants were comparing the strength of the target and the lure to make their confidence judgements. In Study 3b I use a different approach by increasing the presentation frequency of the target and the lure from trial-to-trial. If participants make confidence judgements according to the *relative account* of confidence, they should be insensitive to the increase in strength of the lure. However, if participants use an *absolute strategy* to make confidence judgments, they should be sensitive to the maximum number of presentations of the target.

STUDY 3B.

Methods and Materials

Participants

Fifty people participated through Amazon's Mechanical Turk.

Design

The stimuli for Study 3b were the same as Study 3a. Study 3b was also 1-factor repeated measures design with four levels. The only difference between studies was the presentation frequency of the studied nonsense words. In Study 3a I demonstrated the *relative account* of confidence by making the difference in presentation frequency large between items. For Study 3b I test the *absolute account* of confidence by decreasing the difference in presentation frequency between nonwords and increasing the strength of the target from trial-to-trial. If participants are using the *absolute account* of confidence, then confidence should increase as the strength of the target increases.

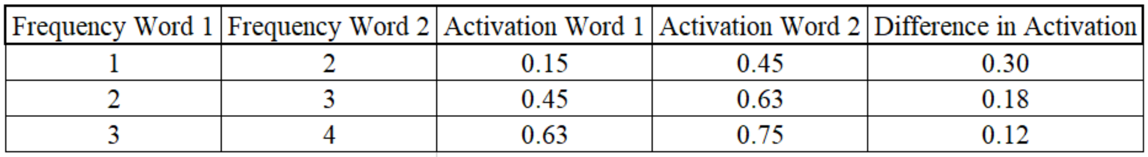
In the study phase, participants could study a nonsense word 1, 2, 3, or 4 times. Participants were tested on nonwords that had been previously rehearsed 0 times (two new words), 1 and 2 times, 2 and 3 times, and 3 and 4 times. Notice that both the rehearsal of the target and the lure increased for each condition, as strength increased for the target so too did the strength for the lure. In this study, an *absolute account* would lead to confidence that increased across conditions, while a *relative account* would lead to a decrease in confidence as strength of the target and lure increased.

Identical to the test phase in Study 3a, nonsense words in the test phase for Study 3b were presented in 32 pairs. The 32 pairs of nonsense words were split into four equal groups based on the frequency of their presentation during the study phase. The first group contained eight pairs of nonsense words that were new and had not been seen during the study phase. The other three groups contained eight pairs of nonsense words studied once and twice (Pairing: 1_2), words studied two and three times (Pairing: 2_3), and words studied three and four times (Pairing: 3_4). Pairing presentations were block randomized such that participants saw only one example of a pairing from each group before seeing another example of a pairing.

We varied the activation for all target trials by presenting the target a different number of times during the study phase (i.e. rehearsals of 2, 3, and 4). Varying the number of presentations for the target from trial-to-trial reduces the relative difference in activation between groups. The goal for Study 3b was to test whether participants would use an *absolute account* of confidence when the difference in activation between nonsense words more difficult to detect.

To be clear, participants can still use the *relative account* of confidence in Study 3b and use the difference in strength between items. Keeping the relative difference of activation constant between pairs of words is very difficult. First, we did not control for the time of the last presentation in the study phase and presentation in the test phase. While this simplification is being addressed in future work, the result is that pairs of words within a group have different absolute and relative activations but a similar relationship between groups. Second, although the difference in the number of rehearsals for each pair of target trials is 1 (i.e. 1_2, 2_3, and 3_4), the difference in activation between pairs gets smaller as target activation increases. **Equation 7** (simplified activation equation) makes the challenge of keeping relative activation constant more apparent: the activation added for each additional rehearsal has diminishing returns The increase in activation for a nonsense word rehearsed 0 times and then 1 time is larger than the increase in activation for a word rehearsed 1 time and then 2 times. Thus, the difference in activation for group 1_2 is larger than the difference in activation for group 3_4. Table 2 shows the activation for each group of nonsense words by group in Study 3b, where n is the number of rehearsals. We use a t value of .5 seconds to make the differences between nonsense words and pairs clearer.

Table 2. Theoretical Activation for Study 3b.



Frequency Word 1	Frequency Word 2	Activation Word 1	Activation Word 2	Difference in Activation
1	2	0.15	0.45	0.30
2	3	0.45	0.63	0.18
3	4	0.63	0.75	0.12

A calculation of activation shows that people have two signals to choose from when making a confidence judgement in Study 3b. If people use the *absolute account* to make their confidence judgement, then confidence for correct trials should increase as the presentation frequency increases (i.e. **Table 2**—Activation Word 2). Alternatively, if people use the *relative account* to make their confidence judgment, then confidence should decrease as the difference in activation decreases (i.e. **Table 2**—Difference in Activation).

Measures

We calculated the mean accuracy between groups of target trials and the mean confidence rating for correct trials only during the test phase.

Results

Fifty participants completed the study. Identical to Study 3a we did not include data from the 0_0 condition. There were 1,200 trials for analysis.

Accuracy

There was a significant difference in accuracy between the 1_2 (M = 69.25, SD = .16), 2_3 (M = 61.00, SD = .16), and 3_4 conditions (M = 57.39, SD = .18), $F(2,98)$ = 7.42, MSE = .19, $p < .05$, η_p^2 = .13. As revealed by a polynomial contrast, as the difference in activation between the target and the lure decreased, the ability to discriminate which word was seen more familiar decreased, Linear: $t = -3.56$, $p < .05$ (**Figure 10**).

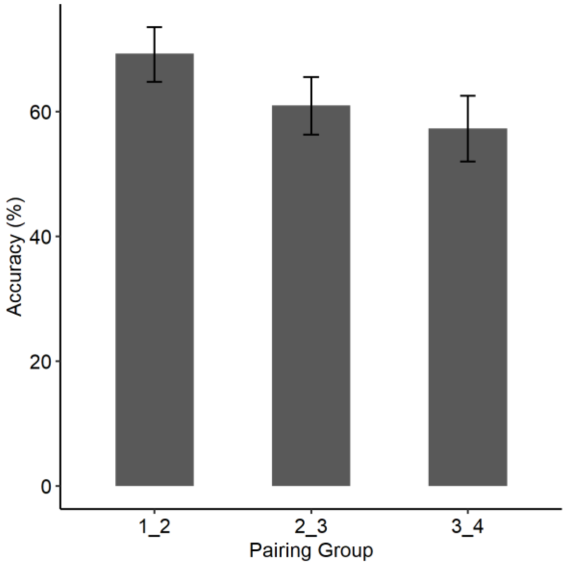


Figure 10. Accuracy for Study 3b. Error bars are 95% confidence intervals.

Confidence

There was a significant difference in confidence between the 1_2 (*M* = 4.82, *SD* = .78), 2_3 (*M* = 5.02, *SD* = .85), and 3_4 conditions (*M* = 5.17, *SD* = .89), *F*(2,98) = 5.37, *MSE* = 1.44, $p < .05$, $\eta_p^2 = .10$. As revealed by a polynomial contrast and seen in **Figure 11**, confidence was sensitive to activation for the target and not the difference in activation between the target and the lure. As activation for the target increased, confidence linearly increased for correct trials, Linear: $t = 2.01$, $p < .05$).

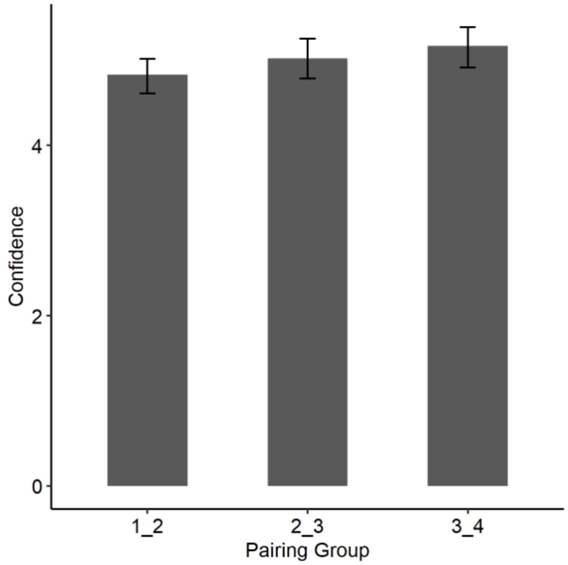


Figure 11. Confidence for correct trials only in Study 3b. Error bars are 95% confidence intervals.

Discussion

In Study 3b, the accuracy data suggests that discriminability decreased as the presentation frequency of the target and the lure increased. The theoretical activation calculated in **Table 2** provides an explanation for this finding. The amount of activation added to base-level activation has diminishing returns with each additional presentation. Therefore, the difference in strength between the target and the lure decreases with additional presentations. Consistent with the *absolute account* of confidence, I found that confidence increased as discriminability decreased. These results suggest that participants used the strength of the target to make their confidence judgment, ignoring the contribution of the lure.

General Discussion

In two experiments I manipulated presentation frequency to investigate the *relative* and *absolute accounts* of confidence judgments in a 2AFC recognition memory paradigm. Use of a recognition memory paradigm allowed me to utilize activation as defined by *ACT-R* to form my predictions. The use of activation is a more precise way to create comparisons that are likely to encourage participants to use one account of confidence over the other. In addition, I used nonsense words so that I could add activation to a stimulus without having to worry about prior experience. Importantly, I used the same stimuli in both studies, only changing the proportion of practice between each word pair.

The *relative account* of confidence suggests that participants use the difference in strength from both stimuli in a 2AFC. As the difference in strength decreases, so will confidence. The *absolute account* makes the opposite prediction. Participants only use the strength for the target to inform their confidence judgment, completely ignoring the contribution of the lure. Although there is some evidence that participants may use the activation from the target and the lure in a summative manner (Zawadzka et al., 2017), the predictions in this study are the same: as the strength for the target increases so should confidence. I found evidence for both the *relative* and *absolute account* of confidence by manipulating absolute and relative activation between groups of nonsense words. The results of this study suggest a strategic component to the formation of confidence judgments.

Study 3a showed a pattern indicative of the *relative account* of confidence: confidence decreased as discriminability between stimuli decreased. Using relative strength to form a confidence judgment seems intuitive but was not a foregone conclusion. The *relative account* of confidence predicts that the strength for each item should be compared. A larger difference in strength (i.e. 0_5) should result in higher confidence than smaller differences in strength (i.e. 4_5). Alternatively, if the representation of confidence in Study 3a was more like the *absolute account*, then confidence in target trials should be the same across groups. The target trials all had the same number of rehearsals (i.e. 5). Because confidence decreased as the difference in strength between items decreased, participants likely used the strength from the item and the lure to make their confidence judgment. In combination with a large body of evidence in support of the *relative account* of confidence for perceptual stimuli (Baranski & Petrusic, 1998; Merkle & Van Zandt, 2006; Pleskac & Busemeyer, 2010; Vickers, 2014), I suggest that the *relative account* of confidence in 2AFC judgments is the default mode of forming a confidence judgment.

However, when the relative strength between items is not available or is weak, participants can use the *absolute account* of evidence. In Study 3b the difference in practice between word pairs was just one repetition making the relative strength of evidence small compared to the relative strength between word pairs in Study 3a. Other investigations of confidence also elude to the *absolute account* being a secondary mode of calculating confidence in 2AFC tasks. In several studies, including this one, confidence increased as discriminability of the items decreased (Maniscalco et al., 2016; Miyoshi et al., 2018; Zylberberg et al., 2012). Maniscalco, Peters, & Lau (2016) describe the *absolute account* in 2AFC tasks as a suboptimal heuristic because confidence has an inverse relationship with the difficulty of the task.

One major contribution of Studies 3a and 3b is evidence of a strategy component to confidence judgments. According to current models of confidence, people should be consistent in how they make a confidence judgment: people either use the difference in strength between items or the strength of the target. Neither the *absolute account* nor the *relative account* of confidence proposes that people can switch to a different strategy. Studies 3a and 3b demonstrate an alternative account: both models of confidence can be demonstrated by manipulating simple features of memory in a subtle manner.

GENERAL DISCUSSION/SYNOPSIS

Many different features of memory have been hypothesized to be the signal used to make memory judgments such as strength, fluency, retrieval time, and frequency. In addition, many different models cite various parameters and mechanisms used to make a confidence judgment. The plethora of features of memory and models used to explain confidence has made the meta-memory and decision-making literature incredibly complex. In this dissertation I attempted to collapse previous work on confidence judgments into two main points.

First, activation as defined by *MFG* is the signal people use to make confidence judgments. In this dissertation I demonstrated that activation is a strong candidate as the source of confidence judgments in memory tasks and not just an in-name-only replacement of previously studied features of memory. Activation has a long history of research and has the benefit of being formally instantiated in an equation which considers the relationship between activation, previous memory retrievals, and time. Furthermore, as described in *MFG,* activation is influenced by primacy, recency, and decay. *MFG* allows for robust predictions to be made for confidence judgments in dynamic decision-making tasks.

For Study 1, I analyzed confidence judgments for sequence errors and demonstrated a unique confidence carryover effect. Confidence was lower for sequence errors that were farther from the next correct step of the task. Current models of confidence such as *2DSD*, *RTCON2*, and the *Poisson model* assume that confidence judgments are made independently and would need additional assumptions to account for my data.

For Study 2, confidence carryover was demonstrated again following the recovery of attention and memory after an interruption. In line with previous work from Altmann & Trafton (2007) I found that nearly half of the participants showed recovery of confidence following an interruption. I formalized the recovery of confidence mathematically using associative strength as a parameter. Associative strength implies that participants benefited from a priming mechanism which is not specified by current models of confidence judgments. As an example, diffusion models such as *2DSD* use drift rate to inform confidence judgments. However, it is not clear why an interruption would reduce drift rate (Buseymeyer, 2018) nor how confidence would increase from trial to trial after an interruption.

Alternatively, *MFG* has a readily available mechanism to explain changes in confidence for dynamic decision-making tasks such as the ones used in Study 1 and Study 2. In Study 1, the interaction of the strengthening constraint, priming constraint, and decay predicts that activation should be lower for sequence errors farther from the next correct step of the task. Confidence followed a similar pattern: confidence was lower for sequence errors farther from the next correct step of the task. For Study 2, nearly half of the participants showed a systematic increase in confidence following an interruption. *MFG* predicts that associative strength forms between trials in procedural tasks like the one used for Study 2. These associative links allow priming to accumulate in uninterrupted portions of the task. To describe the increase in confidence following an interruption, I built a mathematical model that suggests participants utilized a priming mechanism driven by associative strength between trials.

A second contribution of this dissertation was the demonstration of a strategic component of confidence judgments based on two different explanations for how confidence judgments are formed. The *relative account* of confidence suggests that in 2AFC tasks people use the difference in strength between items to inform a confidence judgment. Alternatively, the *absolute account* of confidence suggests that people use the strength of the target in 2AFC tasks and ignore the contribution of the lure. The *relative* and *absolute account* of confidence is based on *sequential sampling models* and *SDT*, respectively. However, neither the *relative account* nor the *absolute account* of confidence has a strategy component where people change how they produce a confidence judgment based on the context of the task. In Study 3a I produced a pattern of confidence judgments in support of the *relative account* and in Study 3b a pattern of confidence judgments in support of the *absolute account*. The results of both studies were accomplished with subtle manipulations of memory. The major difference between each study was the presentation frequency of targets and lures from trial-to-trial. In Study 3a, the difference in strength between targets and lures was large. In Study 3b, the difference in strength between targets and lures was small.

Taken together these three studies demonstrate a need for current models of confidence to account for several new effects. First, current models need to account for how memory manipulations such as interruptions influence confidence. Second, current models need to account for confidence carryover and researchers should expand their models beyond discrete trial paradigms. Finally, current models need to provide a method to switch between the *relative* and *absolute* account of confidence.

MFG provides a simpler explanation for the effects demonstrated here: people use activation to create their confidence judgments. These findings could inform predictions for confidence whenever the best choice for a decision is unclear. For example, the experience that confidence and accuracy are positively related is so pervasive that the U.S. Supreme Court ruled in the case of Neil v. Biggers (1972) that highly confident eyewitness identifications are likely to be accurate (*Neil v. Biggers*, 1972). However, research into eyewitness identification since 1972 suggests that the relationship between confidence and accuracy is more complicated (Brewer & Wells, 2006; Roediger et al., 2012; Sampaio & Brewer, 2009). As a result of some of this research the U.S. Justice Department released guidance in 2017 for eyewitness identification. This guidance suggests that only the first confidence response given by an eyewitness should be used as evidence of an accurate identification (United States Department of Justice, 2017). If confidence is an evaluation of activation, we would expect that multiple instances of eyewitness identification would increase activation through rehearsal, thereby increasing confidence. The policy of taking only the first confidence response in a lineup should limit inflated confidence judgments as a result of repeated identification.

REFERENCES

Aguiar, N., Zish, K., McCurry, J. M., & Trafton, J. G. (2016). Interruptions Reduce Performance across All Levels of Signal Detection When Estimations of Confidence are Highest. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 254–258.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, *26*(1), 39–83.

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review*, *14*(6), 1079–1084.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General*, *143*(1), 215.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341–380.

Ariely, D., Tung Au, W., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, *6*(2), 130.

Baars, B. J. (1992). *Cognition and language. Experimental slips and human error: Exploring the architecture of volition. New York, NY, US*. Plenum Press.

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.

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.

Botvinick, M., & Bylsma, L. (2005). Distraction and action slips in an everyday task: Evidence for a dynamic representation of task context. *Psychonomic Bulletin & Review*, *12*(6), 1011–1017.

Botvinick, M., & Plaut, D. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*(2), 395.

Bradley, J. V. (1968). *Distribution-free statistical tests*.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*(1), 3.

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.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178.

Byrne, M. D., & Davis, E. M. (2006). Task structure and postcompletion error in the execution of a routine procedure. *Human Factors*, *48*(4), 627–638.

Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., & Monk, C. A. (2011). Mitigating disruptive effects of interruptions through training: What needs to be practiced? *Journal of Experimental Psychology: Applied*, *17*(2), 97.

Chung, P. H., & Byrne, M. D. (2008). Cue effectiveness in mitigating postcompletion errors in a routine procedural task. *International Journal of Human-Computer Studies*, *66*(4), 217–232.

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.

Cooper, R. P., & Shallice, T. (2006). *Hierarchical schemas and goals in the control of sequential behavior.*

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17*(4), 297–338.

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions in humans. *ELife, 8*.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778.

Desender, K., Van Opstal, F., & Van den Bussche, E. (2017). Subjective experience of difficulty depends on multiple cues. *Scientific Reports*, *7*(1), 1–14.

DeSoto, K. A., & Roediger, H. L. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science*.

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112–121.

Dougherty, M. R. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General*, *130*(4), 579.

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Ferrell, W. R. (1995). A model for realism of confidence judgments: Implications for underconfidence in sensory discrimination. *Perception & Psychophysics*, *57*(2), 246–254.

Garrett, H. E. (1922). *A study of the relation of accuracy to speed* (Vol. 8). Columbia university.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506.

Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, *50*(4), 243–250.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6*(5), 685–691.

Hiatt, L. M., & Trafton, J. G. (2015). *An activation-based model of routine sequence errors*. NAVAL RESEARCH LAB WASHINGTON DC.

Hintzman, Dougla 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.

Hintzman, Douglas L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411.

Hodgetts, H. M., & Jones, D. M. (2006). Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1120.

Horry, R., & Brewer, N. (2016). How target–lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General*, *145*(12), 1615.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, 391–422.

Johnson, D. M. (1939). Confidence and speed in the two-category judgement (Vol. 34)(No. 241). *Columbia University*.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*(1), 1–24.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. *Neuron*, *84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology*, 329–343.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349.

Li, S. Y., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: An account based on the activation-based goal memory model. *Journal of Experimental Psychology: Applied*, *14*(4), 314.

Maniscalco, B., Peters, M. A., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics*, *78*(3), 923–937.

MCWord: An Orthographic Wordform Database. (n.d.). Retrieved August 14, 2019, from http://www.neuro.mcw.edu/mcword/

Merkle, E. C., & Van Zandt, T. (2006). An application of the poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, *135*(3), 391.

Miyoshi, K., Kuwahara, A., & Kawaguchi, J. (2018). Comparing the confidence calculation rules for forced-choice recognition memory: A winner-takes-all rule wins. *Journal of Memory and Language*, *102*, 142–154.

Miyoshi, K., & Lau, H. (2020). A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychological Review*.

Neil v. Biggers, 409 US. (Supreme Court 1972).

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (Vol. 26, pp. 125–173). Elsevier.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*(1), 1.

Parasuraman, R., & Davies, D. R. (1984). *Varieties of attention* (Vol. 40). Academic Press New York:

Peters, M. A., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., & Halgren, E. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*(7), 0139.

Peterson, H. W., & Birdsall, T. G. (1953). *The theory of signal detectability. University of Michigan: Electronic Defense Group*. Technical Report.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697.

Reason, J. (1990). *Human error*. Cambridge university press.

Roediger, H. L., Wixted, J. H., & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. *Memory and Law*, 84.

Roediger III, H. L., & DeSoto, K. A. (2014). Confidence and memory: Assessing positive and negative correlations. *Memory*, *22*(1), 76–91.

Samaha, J., Iemi, L., & Postle, B. R. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition*, *54*, 47–55.

Sampaio, C., & Brewer, W. F. (2009). The role of unconscious memory errors in judgments of confidence for sentence recognition. *Memory & Cognition*, *37*(2), 158–163.

Tanner Jr, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401.

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications*, *2*(1), 49.

Thomson, R., Harrison, A. M., Trafton, J. G., & Hiatt, L. M. (2017). An account of interference in associative memory: Learning the fan effect. *Topics in Cognitive Science*, *9*(1), 69–82.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, *58*(5), 583–603.

Trafton, J. G., Altmann, E. M., & Ratwani, R. M. (2011). A memory for goals model of sequence errors. *Cognitive Systems Research*, *12*(2), 134–143.

Trafton, J. G., & Monk, C. A. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics*, *3*(1), 111–126.

Treisman, M., & Faulkner, A. (1984). The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 119.

United States Department of Justice. (2017, January 6). *Justice Department Announces Department-Wide Procedures for Eyewitness Identification*. https://www.justice.gov/opa/pr/justice-department-announces-department-wide-procedures-eyewitness-identification

Van Meter, D., & Middleton, D. (1954). Modern statistical approaches to reception in communication theory. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 119–145.

Van Overschelde, J. P. (2008). Metacognition: Knowing about knowing. *Handbook of Metamemory and Memory*, *47*, 71.

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1147.

Vickers, D. (2014). *Decision processes in visual perception*. Academic Press.

Wallsten, T. S., & González-Vallejo, C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review*, *101*(3), 490.

Whittlesea, B. W., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*(1), 62–79.

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489.

Zawadzka, K., Higham, P. A., & Hanczakowski, M. (2017). Confidence in forced-choice recognition: What underlies the ratings? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 552.

Zish, K., Hassanzadeh, S., McCurry, J. M., & Trafton, J. G. (2015). Interruptions can Change the Perceived Relationship between Accuracy and Confidence. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*, 230–234.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.

BIOGRAPHY

Kevin Zish graduated from The College of New Jersey in 2011 with a Bachelor of Arts in Cognitive Science. He moved to Fairfax Virginia in 2012 to begin a graduate program at George Mason University in Human Factors. Kevin received his Master of Arts in Human Factors from George Mason University in 2014.