

University of Arkansas, Fayetteville

**ScholarWorks@UARK**

---

Graduate Theses and Dissertations

---

12-2016

## **Fitting Eyewitness Identification and Confidence to a Diffusion Model of Processing**

Brittany Nicole Race

*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Cognitive Psychology Commons](#), [Criminology Commons](#), and the [Social Psychology Commons](#)

---

### **Citation**

Race, B. N. (2016). Fitting Eyewitness Identification and Confidence to a Diffusion Model of Processing. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/1790>

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu), [uarepos@uark.edu](mailto:uarepos@uark.edu).

# Fitting Eyewitness Identification and Confidence to a Diffusion Model of Processing

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts in Psychology

by

Brittany Race  
University of Arkansas  
Bachelor of Arts in Psychology, 2011

December 2016  
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

---

Dr. James Lampinen  
Thesis Director

---

Dr. Denise Beike  
Committee Member

---

Dr. Bill Levine  
Committee Member

## Abstract

It is necessary to better serve justice to understand the mechanisms behind eyewitness identification and reports of confidence. The material contained within attempt to fit eyewitness identification to a diffusion model of processing, RTCON (Ratcliff & Starns, 2009). Participants saw eight mock crime videos and were then tasked with using eight showups or eight lineups to identify the suspects within the video. Half of the presentations were target present and half were target absent. Additionally, participants were either presented with biased or unbiased instructions. Strangely, unbiased lineups led to higher hit rates which is contrary to most findings in the field. The key elements were comparing the ROC curves of the collected data with the ROC curves of simulated data from the RTCON model. The variables manipulated in the model included  $\mu$ , a variable of memory strength; confidence criteria, a variable of response bias; and decision criteria, a second variable of response bias. The ROC curves derived from the simulated data were poor matches to the collected data. This finding leads to questions on the applicability of this model of diffusion processing to eyewitness research.

## Acknowledgements

The author would like to acknowledge the work and support given from her advisor, Dr. James Lampinen, and her committee members, Dr. Denise Beike and Dr. Bill Levine. She would also like to acknowledge the diligent work of her undergraduate research assistant, without whom, the research could not have been completed.

## Dedication

The author would like to dedicate the following work to her mother, who is always a source of strength and support.

## Table of Contents

Introduction.....	1
Lineups vs Showups.....	2
Instructions for Choice.....	4
Confidence.....	5
Reaction Time.....	7
Diffusion Model.....	8
Hypotheses.....	9
Method.....	10
Participants.....	10
Materials.....	10
Procedure.....	11
Data Simulation.....	12
Results.....	14
Behavioral Results.....	14
Response Time and Confidence.....	16
ROC Analyses.....	18
Simulated Model Results.....	20
Overall Comparisons Between Empirical and Simulated Data.....	32
General Discussion.....	33
Behavioral Data.....	33
ROC Debate.....	34
RTCON Simulations.....	36

Other Confidence Theories.....	38
Conclusions.....	39
References.....	42
Appendix A: Letter of Research Compliance Approval.....	47

## Introduction

More and more convictions are being overturned due to exoneration through DNA testing. A large number of these false convictions have one thing in common, an eyewitness misidentification; the Innocence Project has the percentage of these false convictions is currently reported at 72% (“Eyewitness Misidentification,” n.d.). This circumstance has been replicated in the lab with a study done by Loftus (1974) among others. In this study she asked participants to make judgments of guilt for an armed robbery that resulted in two deaths. Participants either got a small amount of incriminating evidence and no statement of witness identification, a report that a clerk identified the suspect, or a report that the clerk identified the suspect but had poor vision. In the first condition, 18% convicted the defendant; the other two conditions had significantly higher rates (72% and 68%, respectively. As such it is important to understand the possible mechanisms underlying these identifications and what makes them so convincing to a jury.

Before getting into some of the more technical aspects of witness identification, I would like to identify a few cases that underline the possible miscarriages of justice due to faulty identification. One of the most well known is likely that of Ronald Cotton. Cotton was initially accused of raping one woman, Jennifer Thompson, and eventually a second woman. He was identified in open court by Thompson. He was later exonerated through DNA evidence after serving 10.5 years (“Ronald Cotton,” n.d.).

Another incident is that of Willie Williams. He was tried and convicted of raping a woman in Georgia. When she identified Williams in court, she was asked on a scale of 1-100 how sure she was that he was the perpetrator. She responded, “One hundred and twenty” (Rankin, 2008). Based on this identification, Williams served 21 years in prison before being exonerated through DNA evidence.



In both cases the witnesses were highly confident when they identified the supposed perpetrator. Confidence in a selection can be highly persuasive to a jury, even though this confidence is not always justified. The instructions given for identification can also influence not only accuracy, but also the confidence in that choice.

### **Lineups vs Showups**

Lineups are one of the most well-known identification procedures to the public. When a crime show depicts a witness identifying a suspect, it typically occurs in a police station with either a live simultaneous lineup or a six-person picture lineup. This is not how identifications are always made though. Showups include a single suspect that the witness is asked if they recognize (Smith, Bertrand, Lindsay, Kalmel, Grossman, & Provenzano, 2014). A survey of police officers from Canada and America found that over a third of officers had used a showup in the past year for identification (Smith et al., 2014). Smith et al. also examined how often people will say yes to an initial show up. Ninety-three percent of people correctly identified a suspect from an initial showup. Fifty-five percent of people incorrectly identified a foil as the suspect from an initial showup (Smith et al.). This would be equivalent to an officer finding an innocent person in the area who matched a witness's description, showing the witness the innocent person, and having the witness say "yes, that's the person who committed the crime." Since there are no foils in a showup, a witness cannot choose a known innocent foil. Several courts have acknowledged this limitation and define showups as inherently biasing (Bradley v. State, 1980; Stovall v. Denno, 1967).

One of the differences between lineups and showups that may cause issues is the similarity of foils or innocent suspects to the offender (Levi, 1995). The similarity of foils to a target face can change the choosing criterion level (Flowe & Ebbeson, 2007) and can induce

distinctiveness for the target face (Wolgalter, Marwitz, & Leonard, 1992). Additionally, selecting the suspect based on an initial description, but selecting foils by any different method (whether similarity to suspect or similarity to description of the suspect), leads to a bias against the suspect (Navon, 1992). This can be a critical difference as a showup can only use an absolute method of decision making, comparing the presented option to the suspect from memory. Simultaneous lineups, however, allow for a relative process, comparing each lineup option to the rest and choosing the best match to the suspect from memory (Lindsay & Wells, 1985). The mere selection of foils, which would occur in both simultaneous and sequential lineups, causes the lineups to differ from showups. Simultaneous lineups also force a choice between  $>1$  choice, by definition. Whether this choice is identification of a suspect or rejection of a lineup, having to make multiple choices changes behavior (Lindsay et al., 2013). Since both lineups and showups are used in actual police investigations, it is important to understand how they are similar and different to each other and possibly expose any issues

Very little research directly compares simultaneous lineups to showups. Instead, the largest area of research has been comparing simultaneous to sequential (e.g., Carlson, 2011; Mickes, Flowe, & Wixted, 2012; Wells, Steblay, & Dysart, 2014). This displays an obvious hole in the current research. It would be easy to dismiss showups as simplified versions of a sequential lineup, but given the environment that showups occur in this would be a mistake. As we are conducting this study in a lab setting, it is not as possible to induce those circumstances. Through instructions, however, we attempted to create an environment distinct from the lineup condition. A meta-analysis compiled data from eight papers (seven published) that compared showups to either a sequential or simultaneous lineup and found that showups had higher accuracy rates than the lineups (69% compared to 51%). This may be due to a significantly

smaller choosing rate in showups compared to lineups (46% compared to 71%) (Stebly, Dysart, Fulero, & Lindsay, 2003). These differences suggest there may be something wrong with how lineups are being conducted as well.

### **Instructions for Choice**

Ideally, everyone in the world would get essentially the same instructions when asked to identify a suspect from a lineup or a showup, but that just does not happen in actuality. Instructions vary greatly between states and can vary between departments within a state (National Institute of Justice, 2013). In showups, it can be even worse since they typically happen in the field. A biased lineup is typically one where the participant or witness is not reminded that the suspect may not be in the array they are choosing from. An unbiased lineup is one where the participant or witness is explicitly reminded that the suspect may not be in the array and that it is just as important to exonerate an innocent suspect as it is to identify a guilty one.

One study found that participants were 3.2 times more likely to choose an innocent suspect with biased instructions compared to unbiased instructions (Quinlivan et al., 2012). Examining instruction bias in a cue-belief model allows one to better understand some of the effects on confidence and accuracy. Specifically biased lineup instructions can represent an incongruent cue, which can inflate confidence while deflating accuracy (Leippe, Eisenstadt, & Rauch, 2009). Part of this decrease in accuracy is due to lower rejection rates. Biased instructions have been found to lower lineup rejection rates for both a suspect lineup and a witness lineup (Brewer & Wells, 2006). A meta-analysis of experimental lineup procedures found a significant decrease in accuracy for target-absent lineups when biased instructions were used, but inconsistent results for target present lineups (Clark, 2005). The big difference here is

that biased lineup instructions lead to higher false positives compared to unbiased lineup instructions (Clark, 2005; Steblay, 1997). This has been found for both child witness and adult witnesses (Pozzulo & Dempsey, 2006). This choice bias remained true regardless of the potential consequences of their identification, although this was moderated by sex of the participant (Foster, Libkuman, Schooler, & Loftus, 1994).

There does appear to be an asymmetry regarding biased instructions. The increase in false positives has been repeatedly demonstrated to either only exist or be much stronger in target-absent lineups compared to target-present (Paley & Geiselman, 1989; Malpass & Devine, 1981). Problematically, attorneys did not find biased lineup instructions to be significantly less fair than unbiased lineup instructions (Stinson, Devenport, Cutler, & Kravitz, 1996). If the above issues hold true, attorneys should be the first to notice the unfairness of lineup instructions. If they cannot see the issues behind biased lineup instructions, how are juries supposed to understand the effects that instructions for choice can have?

### **Confidence**

Accuracy is often the only dependent measure in eyewitness identification or at least treated as the most important one; however, there are other measures that can inform both the process of identification and accuracy for identification. One of these measures is confidence. In a survey of jury-eligible adults, over half thought that confidence was positively correlated with eyewitness accuracy (Brigham & Bothwell, 1983). Similarly, in a mock jury situation, the confidence of a witness significantly correlated with juror-rated believability, but was not significantly correlated with accuracy (Wells, Lindsay, & Ferguson, 1979). After three months, memory for an armed robbery was tested and confidence was only moderately correlated with accuracy (.38). The researchers actually found that emotional impact was more highly correlated

with accuracy and post-event thinking was highly correlated with confidence (Odinot, Wolters, & van Koppen, 2009). Confidence is also used to differentiate truthful statements from deceptive statements. Even when warned not to use confidence as a way to differentiate, participants rated confident false statements as more believable than nonconfident true statements (Tetterton & Warren, 2005).

In addition to the scientific research, confidence has an informing role in evidence inclusion in courts. In *Neil v. Biggers* (1972), the factors to consider for witness inclusion were stated to include the certainty of the witness. *Manson v. Braithwaite* (1977) upheld certainty as a measure for inclusion of witness testimony. There is sometimes explicit confirming feedback in a lineup identification (e.g., the investigator says “you got the guy”), but when a crime goes to trial there is implicit confirmation because the suspect chosen is the one under arrest and being tried. The issue here is that confirming feedback inflates confidence, which as mentioned above can be very persuasive to a jury (Semmler, Brewer, & Wells, 2004). Even when presented evidence that a witness had inflated confidence, there was only a significant difference in juror-believability if the initial confidence statement was video-taped. If the change in confidence was read or asked at trial, the mock jurors did not change their opinions of guilt, sentencing, or credibility between confidence consistent witnesses and confidence inflated witnesses (Douglass & Jones, 2013). This behavior, however, does not appear to be tied to jurors. Law enforcement viewed a video of a witness recounting a robbery. The officers rated confident witnesses as more accurate and more reliable than unconfident witnesses (McClure, Myers, & Keefauver, 2013). Seventy-five percent of prosecutors also think confident witnesses are more accurate. In the same study, only forty percent of defense attorneys think confident witnesses are more accurate (Brigham & Wolfskiel, 1983). A separate study found sixty-four percent of lawyers

think more confidence leads to more accuracy (Rahaim & Brodsky, 1982). Even if there is the moderate correlation between accuracy and confidence, as found in Odinet et al.'s study (2009), there seems to be a widespread overreliance on confidence in determining accuracy or truthfulness, regardless of the judging source.

### **Reaction Time**

A third measure that should be examined when eyewitness identification is examined is reaction time. There can be important information offered by the speed at which a witness chooses. For instance, it has been demonstrated that accurate nonchoosers (8.3 seconds) are significantly faster in their nonchoosing than inaccurate nonchoosers (11.1 seconds) (Sauerland, Sagana, & Sporer, 2012). Similarly, choosers' accuracy is negatively correlated with decision time, meaning the faster they chose the more likely they were accurate (Smith, Lindsay, Pryke, & Dysart, 2001). Furthermore, when responses were split into less than 15 seconds and greater than 15 seconds, there was a significant difference in accuracy rates, 63% and 35% respectively. This stayed true regardless of whether the suspect was from the choosers' own race or opposite race. This same pattern was seen in a similar study that split decision making times into 1-15, 16-30, and >30 seconds. The shortest decision time was the more accurate than either of the other two; the middle time segment was more accurate than the longest time segment (Smith, Lindsay, & Pryke, 2000). These studies suggest a fairly strong association between decision time and accuracy; however, there has been evidence that shows this negative relationship only holds true for choosers. One experiment found that for choosers there was a significant negative correlation between accuracy and decision time, but for nonchoosers there was a nonsignificant positive correlation between accuracy and decision time (Sporer, 1993).

## Diffusion Model

Additionally, we will attempt to fit eyewitness identifications to a diffusion model of processing using Ratcliff and Starns (2009) RTCON model. There are specific assumptions for this model including: the information available for comparison is not a single score, but a normal distribution of evidence; the distribution is divided into confidence intervals or accumulators; each accumulator has its own boundary; as one accumulator is increased, others decrease to an equal amount; and the evidence cannot fall below zero. At every point in the decision making process ( $\Delta t$ ), an accumulator is randomly chosen and receives a change ( $\Delta x(t)$ ) determined by the evidence available in that region or the drift rate ( $v$ ) plus an amount of noise, which represents the variability necessary in processing. Abiding by the previous assumption, for each change, an equal change in the other accumulators occurs.

There are several particular parameters necessary for the model. The first is a scaling parameter, which applies to the drift rate. This initial parameter can be thought of as analogous to the general speed through which evidence accumulates. It affects another parameter, the drift rate. The drift rate is in relation to the match between memory and stimulus and so can be thought of as an element of memory strength. The shorter the drift rate, the stronger the memory. A second parameter,  $\mu$ , is another parameter relating to memory strength. Specifically, a high  $\mu$  means a strong across trial memory match. The third is time taken up with other processes. The fourth is variability in nondecision time. The fifth is a standard deviation in evidence accumulation. The sixth required parameter is a within-trial variability in the process. The total number of confidence divisions is equal to a number of additional confidence criterion parameters. The decision criteria make up the remaining parameters with each confidence

criteria existing in both the confirming and disconfirming decision aspect. These final two types of parameters are analogous to response bias.

This model and its initial form has been mostly used for basic perceptual and cognitive tasks (Ratcliff & Starns, 2009; Ratcliff, McKoon, & Tindall, 1994) and to our knowledge has not been applied to something as intricate or with as many real world implications as eyewitness identification.

## **Hypotheses**

For this study, participants participated in either eight lineups or eight showups, half of which were target present and the other half were target absent. Additionally participants were either presented with biased instructions or unbiased instructions. In addition to measures of accuracy, measurements of response times were also collected.

We expect for showups to have higher false alarm rates than lineups, and as a result, be less accurate. We also expect for biased instructions to have higher false alarm rates than unbiased instructions and be less accurate. We also expect for biased instructions to result in significantly higher reported confidence levels. Finally, we have no idea if the research will conform to Ratcliff and Starns (2009) RTCON model. If eyewitness identification fits the model, showups would decrease drift rate as there is less information to sift through in order to make a decision. It is also likely that showups would change the confidence parameters. Since there are no foils, there is no interference that can make one doubt. In the real world, the environment that showups take place in can have influence in reporting confidence. Specifically, showups typically happen in a geographic area very close to the scene of a crime and temporally very close to the time of the crime. This is because showups typically occur when the police find someone in the area who seems to match the description of a suspect and bring them for



identification to the witness. This environment almost certainly has some sort of influence on choosing behavior and confidence reports, but since this study is conducted in the lab, it is premature to make judgments based on this external environment. Biased instructions will increase the confidence parameters at the higher ratings of confidence as participants should assume that the suspect is in the lineup regardless as to whether they actually are. Target-absent lineups should increase the drift rate as the presented face(s) will necessarily not match the memory of the crime. A target-absent lineup would have the highest drift rate according to this model.

## **Method**

### **Participants**

518 participants (320 identified as female) were drawn from the University of Arkansas undergraduate general psychology population. Average age was 20.0 (SD = 5.33) and 79.6% identified as Caucasian, 6.7% identified as Hispanic or Latino, 5.9% identified as Black or African-American, 3.4% identified as Asian, 1.5% identified as American Indian or Alaska Native, 1.3% identified as other, 1.1% declined to respond, and 0.4% identified as Native Hawaiian or Other Pacific Islander.

### **Materials**

All experiments were run using Superlab. For the exposure to the crime, eight scenarios were recorded from 10' away using an iPhone5 camera. Each video lasted approximately 20 seconds and the face of the suspect is directly to camera for eight seconds. The eight scenarios included a lab room theft, a laptop theft, a bike theft, a textbook theft, a backpack theft, mail theft, breaking into a car, and breaking into a house. Half the scenarios took place outside. Half

the scenarios have male suspects. Six person lineups were constructed using description matched foils which were normed for similarity to the description.

## **Procedure**

Participants came in to the lab five at a time. After filling out the informed consent, participants were asked to keep their hand poised over the number key pad throughout the experiment. Each participant saw all eight scenarios. Following each scenario a manipulation check was done to make sure that the participant was attending to the video. Following the manipulation check, participants were immediately presented with either the lineup or the showup for that video. Half of the participants received the six person lineups and half of the participants received show ups. Additionally, half of the participants received biased lineup instructions (not mentioning the possibility of the suspect not being displayed) and half of the participants received unbiased lineup instructions (specifically mentioning the possibility that the suspect may not be displayed) on a slide separate from the suspect slide. Half of the scenarios were target-present lineups/showups and half were target-absent lineups/showups. In the lineup condition, the six photos were randomly assigned to different positions using the randomizer on superlab. The participants made their choice by pushing the number associated with the picture they believed to be the suspect or the seven for “not here.” In the showup, target-absent condition, one photo from the lineup foils was randomly selected and displayed for the participant. The participants pushed one for yes and two for no. Reaction time data was collected for suspect choice. After making a choice, the participants were asked to rate their confidence in their choice on a 0-9 scale (0-10%, 11-20%, 21-30%, etc.). Reaction time data was also collected for confidence rating.

Following the final confidence rating, participants completed a manipulation check to ensure that they read and understood the lineup/showup instructions. They were asked to write down, word for word to the best of their ability, what the instructions were. This was to ensure that the participants read the biased instructions differently from the unbiased.

### **Data Simulation**

The initial plan was to use a collaborator's expertise to train me in data simulation for this type of scenario. Due to unfortunate circumstances, this was not feasible. Instead, with the help of my advisor and a web simulation Ratcliff built and published, we attempted to simulate how the RTCON model could apply to a six-person lineup with the option to reject. I created an Excel spreadsheet with a random number between 0-1 was chosen for 10,000 simulated subjects. In the case of showups, only one random number was generated. For lineups, six random numbers were generated with the first belonging to a simulated suspect.

The Ratcliff website calculates proportions for the different confidence categories ranging from 1 (very sure new) to 6 (very sure old) for both target and foil scenarios. The cumulative proportions were used as headers in the Excel file to determine the simulated decision and confidence response. For showups, these proportions were each listed once to simulated 10,000 target-present showups and 10,000 target-absent showups. The random number generated would fall between one of the six decision criteria. If it was the first, second, or third criteria, it was coded as a rejection of the show up. If the number fell between the fourth, fifth, or sixth criteria, it was coded as an acceptance of the lineup. Categories 1 and 6 were the strongest levels of confidence. Categories 2 and 5 were mid-levels of confidence. Categories 3 and 4 were low levels of confidence.

For lineups, the simulation was a bit more complicated. To simulate a target-present lineup, the target proportions were copied once and the simulated suspect's random number was compared to the cumulative proportions to get a decision and confidence in that decision. The foil cumulative proportions were then copied five times with the respective random numbers compared to those proportions. This led to six different decision/confidence combinations. The final choice for that simulated participant was chosen by the strongest simulated confidence level. If there was a tie, excel randomly chose a winner from those that were tied. If neither the target, nor any of the foils landed in the fourth, fifth, or sixth decision criteria, it was considered a rejection. A similar procedure was used to simulate target-absent lineups. Specifically, rather than having the target proportions copied once and five copies of the foil proportions, there were six copies of the foil proportions, and none of the six random numbers generated acted as a stand in for a simulated suspect. The rest of the procedure was identical to the target-present simulation. Unfortunately, we were unable to come up with a system to model response times, so they are not included in any descriptions of the simulated data.

To examine the parameters that the RTCON model suggests, we systematically changed some of the variables using the website built by Ratcliff. For each parameter, I brought each statistic to the reported base from Ratcliff and Starns (2009) before changing three different parameters based on the restrictions of the specific parameter. The first variable was  $\mu$ , the variable that related to memory strength. According to the model, it cannot go below zero and it began at 0 in the original article. I manipulated the variable by changing it to .5, .75, 1, 1.25, and 1.5, which would be analogous to steadily increasing match between stimulus and memory, or a steady increase in memory strength.

The second set of variables manipulated were those relating to the confidence parameters. The RTCON website uses five confidence parameters. An increase in the confidence parameters shifts the divisions along the x-axis. Because it is assumed that the information is normally distributed, an increase in the confidence parameters also signifies a stronger likelihood for a high level of confidence to be reported. A decrease in confidence would lead to the opposite effects. I manipulated the variables by adding and subtracting .25, .5, .75, and 1 to the values reported by Ratcliff and Starns (2009) across all confidence parameters, leading to eight simulations for the confidence parameters.

The third and final set of variables manipulated were those relating to the decision parameters. The RTCON website uses six decision parameters. One of the assumptions of the model says that an increase in one decision parameter must cause an equal decrease in the other parameters. In order to abide by this parameter as I increased the yes decision parameters, I decreased the no parameters in equal measure. This would represent requiring more evidence to say yes and less evidence to say no, or creating a more conservative response bias. I manipulated the variables by adding .25, .5, .75, and 1 to the values reported by Ratcliff and Starns (2009) to the yes parameter (and subtracting them from the no parameter) and then subtracting .25, .5, .75, and 1 from the values reported by Ratcliff and Starns (2009) to the yes parameter (and adding them to the no parameter).

## **Results**

### **Behavioral Results**

I first analyzed the effects of instruction (Biased vs. Unbiased) and presentation type (Lineup vs Showup) on the hit rate, false alarm rate,  $d'$ , and beta. I also calculated diagnosticity for each of the four conditions. Diagnosticity refers to the odds of guilt given that a suspect has

been identified assuming a prior probability of guilt of 50%. Descriptive statistics for these measures are presented in Table 1.

Table 1  
*Means for Variables Collapsed Across Trials*

		Hit Rate	False Alarm Rate	d'	Beta
Lineups	Unbiased	.71 (.24)	.11 (.05)	.04 (.81)	1.27 (.61)
	Biased	.66 (.25)	.09 (.05)	.30 (.81)	1.43 (.58)
Show up	Unbiased	.84 (.19)	.16 (.22)	1.65 (.69)	1.05 (.55)
	Biased	.83 (.21)	.14 (.18)	1.65 (.67)	1.03 (.51)

*Note.* Standard deviations in parentheses. False alarm rate was divided by six in order to simulate an innocent suspect choice for lineups.

To examine the results of manipulating presentation type and instruction bias on participant accuracy, a MANOVA was conducted using instruction bias and presentation type as independent variables and collapsing the eight trials each participant completed to use signal detection measures as dependent variables, as follows, hit rate, false alarm rates,  $d'$ , and beta.

There was a significant main effect of presentation on hit rate,  $F(1, 504) = 44.87, p < .001$ , false alarm rate,  $F(1, 504) = 6.91, p = .009$ ,  $d'$ ,  $F(1, 504) = 538.63, p < .001$ , beta,  $F(1, 504) = 40.02, p < .001$ . Showup presentation had higher hit rates, false alarm rates, and  $d'$ . Lineup presentation had higher betas. Instruction manipulation approached significance for hit rates,  $F(1, 504) = 3.29, p = .070$ . Bias in instruction had no other significant main effects,  $p$ 's  $> .119$ . Biased instructions had higher hit rates. There was a significant interaction effect between lineup presentation and instruction manipulation for  $d'$ ,  $F(1, 504) = 4.45, p = .035$ , and beta approached significance,  $F(1, 504) = 2.72, p = .099$ . There was a crossover interaction effect for  $d'$  with unbiased lineup instructions having a higher  $d'$  average than biased lineup instructions

and unbiased showup instruction having a lower  $d'$  average than biased showup instructions.

There were no other significant interaction effects,  $ps > .401$ .

Diagnosticity ratios give an idea of how accurate a variable is. It is calculated as the odds ratio of the proportion of guilty suspects identified divided by the proportion of innocent suspects identified.

$$D = p(\text{Suspect ID}|\text{Suspect Guilty})/p(\text{Suspect ID}|\text{Suspect Innocent})$$

For a show up, all mistaken identifications are identifications of an innocent suspect. For a lineup, not all mistaken identifications are identifications of innocent suspects, some are mistaken identification of fillers. For this reason, when calculating diagnosticity for lineups, a common practice is to divide the total proportion of mistaken identifications in target absent lineups by the total number of pictures in the lineup (in the present case six) in order to obtain an estimate of  $p(\text{Suspect ID}|\text{Suspect Innocent})$ .

In this case the diagnosticity ratio for lineups is 6.90 and showups is 5.53.<sup>1</sup> This suggests that if a suspect is identified by a witness the suspect is more likely to be guilty if the witness saw a lineup than if the witness saw a showup. Biased instructions have a ratio of 6.31 and unbiased instructions is 6.90. This suggests that if a witness identifies a suspect, the suspect is more likely to be guilty if the lineup instructions were unbiased than if they were biased.

### **Response Time and Confidence**

Additional behavioral measures included confidence and reaction time for choice and reaction time for confidence. These were collapsed across participants and are shown in Table 2.

---

<sup>1</sup> Diagnosticity was calculated using the raw data which does not have equal unbiased/biased cases. This means the numbers reported differ slightly from using the averages in table 1.

Table 2  
*Means for Additional Behavioral Measures*

		Confidence	Response Time for Confidence	Response Time for Choice
Lineups	Unbiased	.6.39 (2.15)	2987.14 (2314.95)	8843.81 (5317.99)
	Biased	6.33 (2.21)	2952.40 (2347.89)	8897.33 (5056.65)
Show up	Unbiased	7.36 (1.82)	3050.07 (2389.71)	3410.23 (2766.95)
	Biased	7.53 (1.72)	3181.00 (2770.47)	3338.28 (2612.85)

*Note.* Standard deviations in parentheses.

A second MANOVA was used with confidence selection, response time for choice, and response time for confidence selection as dependent variables (means and standard deviations can be found in Table 1.). There were some multivariate outliers and 2.18% trials were removed from this analysis using Mahalanobis Distance. Additionally, each dependent variable was a non-normal distribution and was transformed. Confidence selection and response time for choice underwent a square root transformation. Response time for confidence underwent log transformation. No other assumptions were violated.

There was a significant main effect of presentation on confidence selection,  $F(1, 3994) = 276.39, p < .001$ , and response time for choice,  $F(1, 3994) = 2776.13, p < .001$ , but no significant effect on response time for confidence,  $F(1, 3994) = .059, p = .809$ . Showup presentations had higher confidence selections and shorter response time for choice. There was no main effect of bias in instruction on any dependent variable,  $p > .4$ . There was a significant interaction between presentation and bias in instruction for confidence,  $F(1, 3994) = 5.42, p = .02$ . There was a crossover interaction effect with lineup presentations increasing in confidence from biased instructions (6.37) to unbiased (6.47) and showup presentations decreasing in confidence from

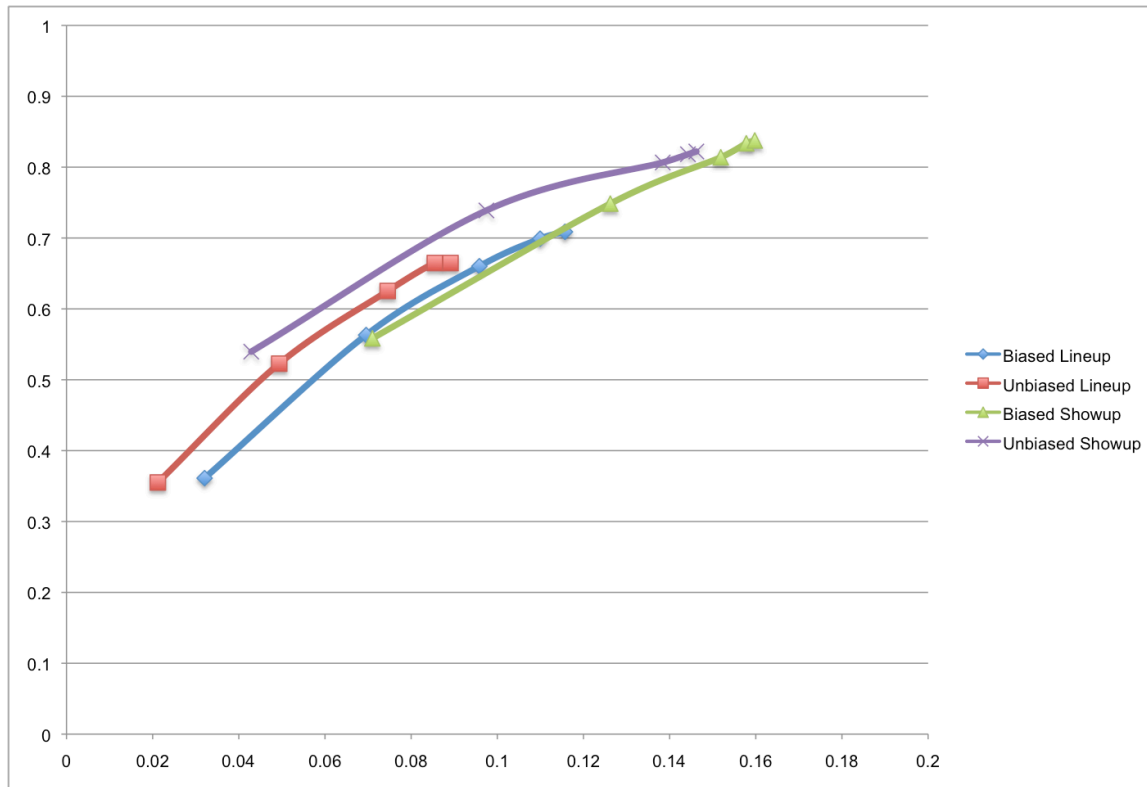


biased instructions (7.58) to unbiased (7.38). There was no significant interaction for either response time variable,  $p$ 's > .2.

## ROC Analyses

**Empirical ROC results.** I next turn to the analysis of ROC results. ROC curves plot cumulative hits against cumulative false alarms at different levels of confidence. A strength of using ROC analyses is that it allows two or more conditions to be compared by looking at the hit rate while equating those conditions on false alarm rate. Recently, it has been argued that ROC curves provide the best analytical methodology for comparing lineup procedures in terms of their practical utility (Wixted & Mickes, 2015). However, critics argue that ROC curves are either ill suited for analyzing eyewitness data or that the advantages of the ROC approach for lineup research have been overstated (Wells, Smith, & Smalarz, 2015).

Although I will touch on this controversy later, the purpose of my research is not to use ROC analyses to sort good procedures from poor procedures, but rather to evaluate underlying theories of the memory and decision processes that might be involved in making lineup/showup decisions. Specifically, the purpose of the present thesis is to examine the ability of the RTCON model (Ratcliff & Starns, 2009) to account for performance in lineup and showup tasks. The model has two specific parameters that represent a change in response bias and one parameter that represents memory strength. The empirical data collected included a manipulation known to change response bias (biased or unbiased instruction), but does not affect memory strength. Figure 1 shows the ROC curves for each of the four cells created by factorially crossing instruction type (biased vs. unbiased) and presentation type (lineup vs. showup).



*Figure 1. ROC Curves by Condition*

Before moving to comparing the empirical ROC results to the simulated ones, there are a couple of noteworthy findings. First, the unbiased showup ROC curve is above (i.e., dominates) the unbiased lineup ROC curve suggesting better performance in showups than in lineups. This finding runs counter to recent findings comparing lineups to show ups using ROC analyses. Gronlund et al. (2012) found drastically different shapes for the ROC compared to this data, and found lineups to have a larger partial area under the curve (pAUC<sup>2</sup>). Wetmore et al. (2015) found a similar trend. Second, showup ROC curves are shifted to the right relative to lineup ROC curves, suggesting that showups resulted in greater overall rates of choosing the suspect, whether guilty or not. This has sometimes been referred to as a difference in response bias, but as Lampinen (2016) pointed out, calling this difference response bias is a somewhat loaded term.

<sup>2</sup> pAUC refers to the space under a ROC curve whose area is positively related to memory strength and discrimination.

One reason that the showup ROC curve may be shifted to the right relative to lineup ROC curves is simply that when one makes a mistake in a target absent lineup, at least some of the time one will select a filler rather than the innocent suspect. Wells, Smalarz, and Smith (2015) called this filler siphoning. Additionally, the biased lineup and biased showup are dominated by the unbiased versions but do not dominate each other. Rather, the biased showup is merely shifted to the right somewhat relative to the biased lineup.

One of the possible reasons for the difference from previous studies is that the participants in this study seemed to have better overall memory. Gronlund et al. (2012) had a false alarm rate for showups of .24 and Wetmore et al. (2015) had a showup false alarm rate of .49 or .40 depending on condition. Both of these results are much larger than those found in the current study. There may also be something of a practice effect showing here. The previously cited studies had each participant only go through one trial. In this experiment, each participant underwent eight randomized trials.

### **Simulated Model Results**

To examine how this data compares to predictions of the RTCON model, I conducted simulations of the RTCON model corresponding to the factors manipulated in my experiment. Each went through a lineup simulation, a showup simulation, and parameter changes meant to mimic the effect of the instruction manipulation.

**Changes in memory strength.** In the original RTCON model, the memory strength parameter is  $\mu$ . Higher  $\mu$ s mean a stronger match between study and test items. To examine the influence of changes in memory strength on the respective dependent variables, I used the parameter estimate for  $\mu$  obtained by Ratcliff and Starns (2009) in their experiment, and then

systematically added to it in increments of .25. This value of change was selected so as to cover a reasonable expanse of the parameter space.

**Lineup.** To model lineups using RTCON, I used the approach used by Lampinen (2016) and Wixted, Mickes, Dunn, Clark, and Wells (2016) in simulating signal detection models of lineups. The model assumed that memory strength is stronger for guilty suspects than for innocent suspects or fillers. It then assumes that each member of the lineup is evaluated against a criterion. If none of the items exceeds the criterion, the lineup is rejected. If one or more items exceeds the minimum threshold, then the item that would produce the highest confidence is chosen. If confidence is equal for two or more items, the winner is chosen randomly. Confidence is assumed to be based on the confidence level assigned to whichever item is selected. These simulated data can then be used to generate hypothetical ROC curves.

The model simulation ROC results can only be looked at qualitatively, but even a qualitative examination of the results seems to lead to a specific conclusion. Hit rates move consistently up with large increases in memory strength, without false alarm moving much at all. This increase in hit rates without a corresponding increase in false alarms does lead to better  $d'$ 's. Beta changes in a small, nonlinear way. Memory strength lines do not differ much in curve, and only have large movement up the y-axis. This is in line with the current thinking of ROCs (see Figure 2).

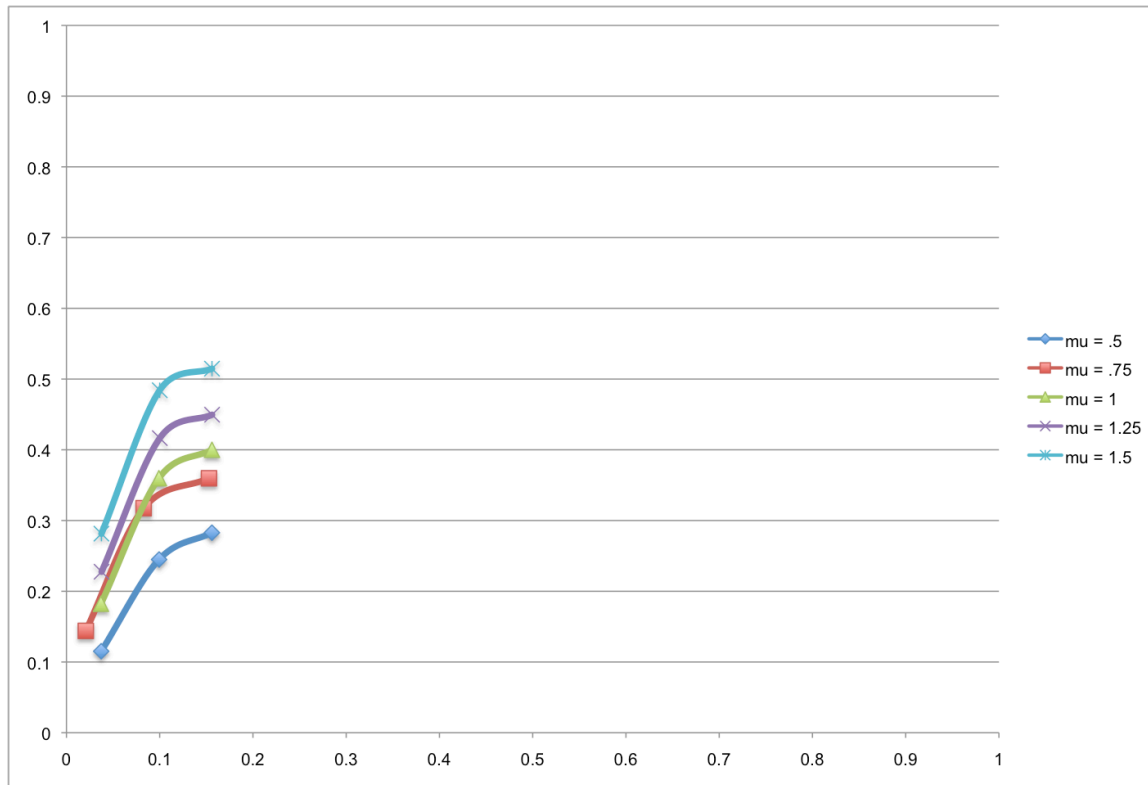
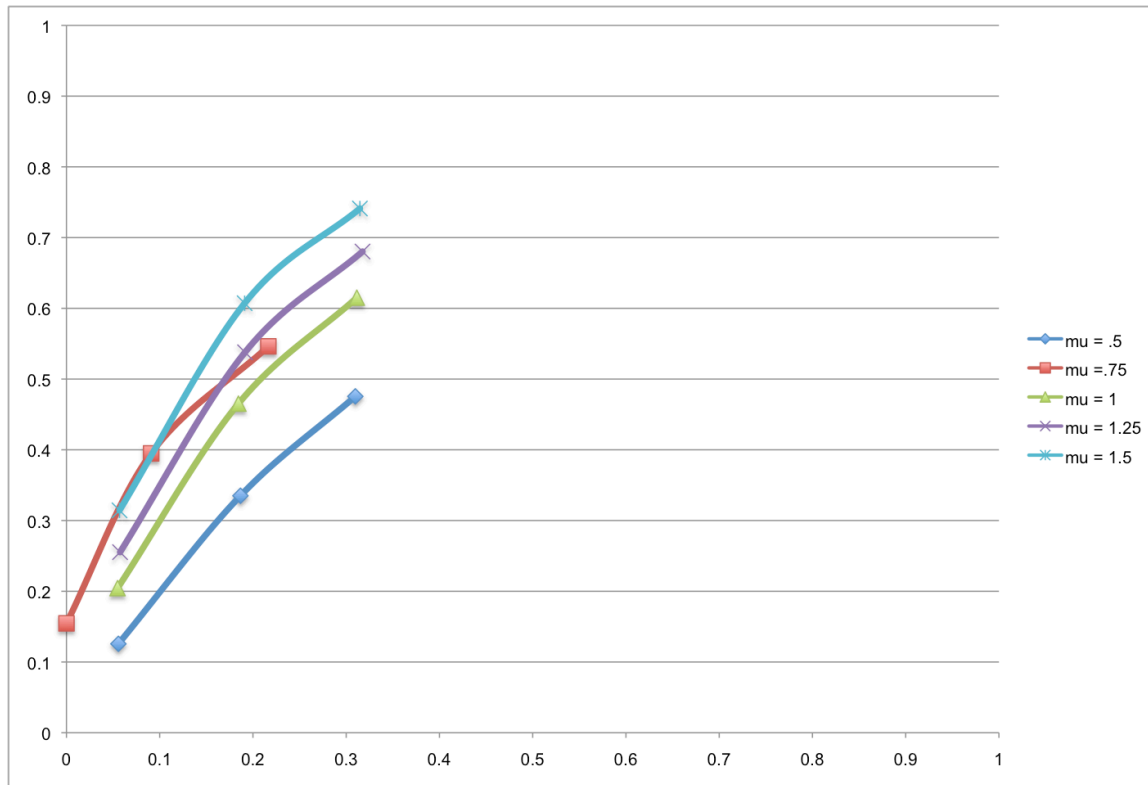


Figure 2. ROC Curves of Simulated Lineup Data with Memory Strength Manipulation

**Showup.** For the showup data, the model simulation changed in similar patterns to the lineup condition. When manipulating the memory strength parameter, hit rates move consistently up with increases in memory strength, without false alarm moving much at all. This increase in hit rates without a corresponding increase in false alarms does lead to slightly better  $d'$ 's. Beta generally decreases, but not in a particularly large way with increases in memory strength. The simulated ROC lines do not differ much in curve, and only have large movement up the y-axis (see Figure 3). Again, this is consistent with the current thinking of ROCs.

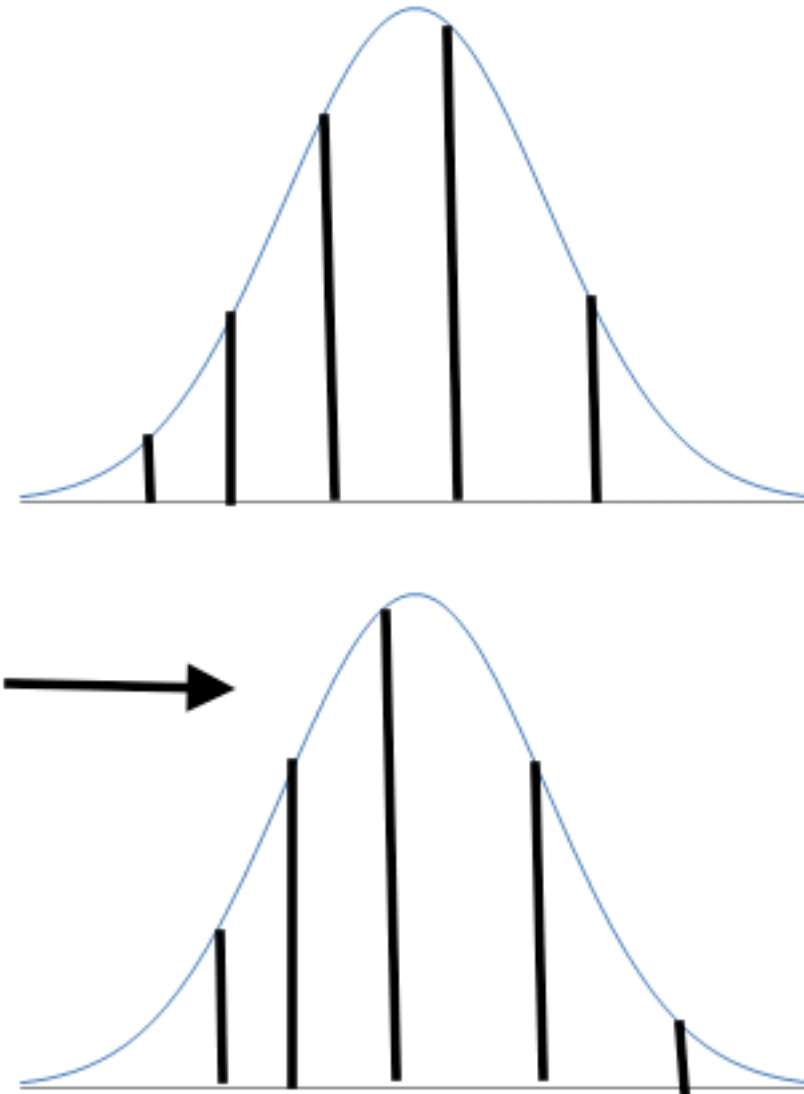


*Figure 3. ROC Curves of Simulated Showup Data with Memory Strength Manipulation*

When manipulating the memory strength parameter which theoretically should not act differently for biased or unbiased instructions, hit rates move consistently up with increases in memory strength, without false alarm moving much at all. This increase in hit rates without a corresponding increase in false alarms does lead to slightly better  $d$ 's. Beta shows almost no movement. Memory strength ROC lines do not differ much in curve, with the peculiar exception of  $\mu = .75$  (see Figure 3). Disregarding that anomaly, there is a consistent movement up the y-axis with no change across the x-axis.

**Changes in confidence criterion.** There are two different parameters in the RTCON model that simulate response criterion. The first is confidence criterion. The basic premise of this parameter is that a normal distribution is divided into different segments of confidence. By

adding to those parameters, those segments shift right, making it easier for higher levels of confidence to be selected (see Figure 4).



*Figure 4. Example of a Shift in Confidence Criterion*

As a reminder, the confidence criteria were manipulated by adding or subtracting across all criteria levels. Changes were made in .25 increments up to 1.

**Lineup.** Increasing the confidence criteria parameters led to lower choosing rates, with lower hits, lower false alarms, little change in  $d'$ , and smaller betas, as seen in Table 3.

Table 3.  
*Signal Detection Variables for Simulated and Real Lineup Data Sets*

	Hit Rate	False Alarm Rate	d'	Beta
Collected Data	.69	.10	.15	1.18
Ratcliff & Starns (2009)	.42	.16	-1.73	3.13
+.25 to confidence	.37	.15	-1.76	2.59
+.5 to confidence	.31	.15	-1.83	2.17
+.75 to confidence	.27	.15	-1.88	1.86
+1 to confidence	.22	.15	-2.02	1.58
+.25 to decision	.30	.15	-1.79	1.95
+.5 to decision	.17	.15	-2.11	1.27
+.75 to decision	.09	.15	-2.48	0.76
+1 to decision	.04	.15	-2.86	0.42
-.25 to confidence	.45	.16	-1.78	3.86
-.5 to confidence	.50	.16	-1.78	4.83
-.75 to confidence	.53	.16	-1.87	6.65
-1 to confidence	.57	.16	-1.98	10.14
-.25 to decision	.46	.16	-2.01	6.24
-.5 to decision	.47	.17	-2.56	21.49
-.75 to decision	.46	.17	-2.42	15.04
-1 to decision	.46	.17	-3.83	1002.12

*Note.* “to confidence” refers to a change in the confidence parameters in the model. “to decision” refers to a change in the decision parameter, with a – referring to a decrease in the yes decision bins and an increase in the no decision bins and vice versa.



The ROC curve for the confidence criterion changes slightly altered the shape of the line from the lowest requirement for confidence to the highest, but the primary change was the line moving up the y-axis. This leads to larger pAUC (see Figure 5).

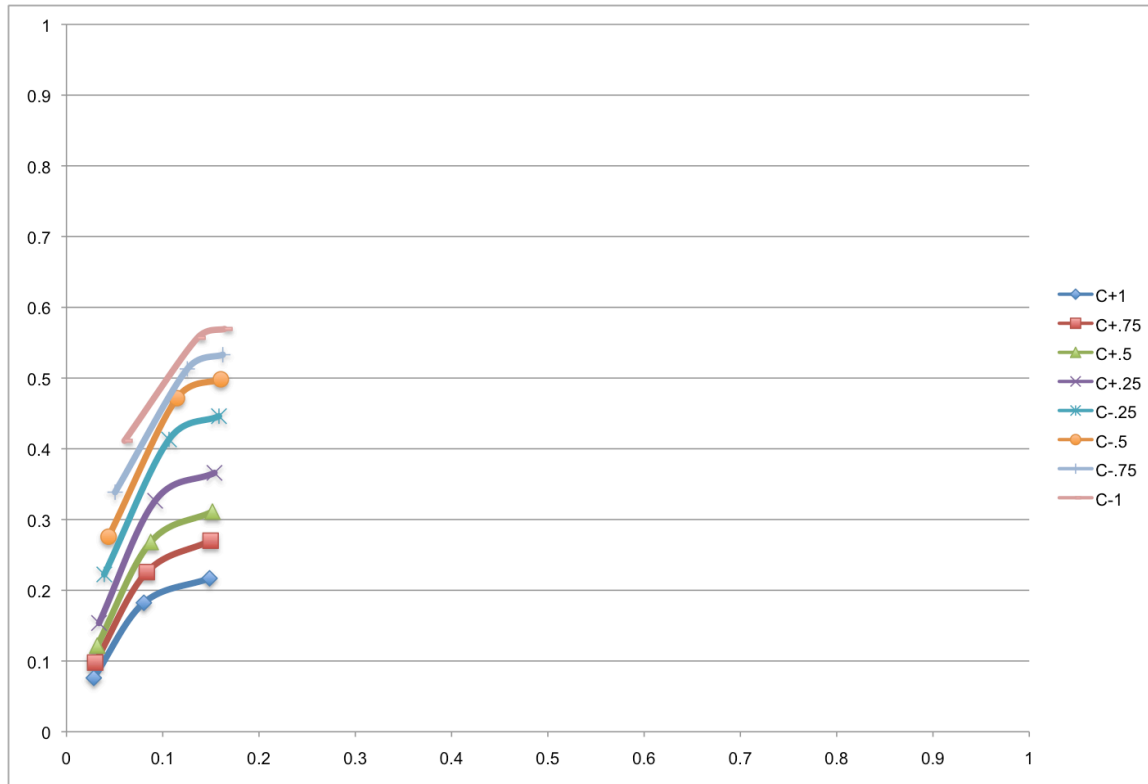


Figure 5. ROC Curves of Simulated Lineup Data with Confidence Criterion Manipulation

This flies in the face of what is typically thought regarding ROC curves according to standard signal detection theory. Theoretically only aspects that manipulate memory strength should change the pAUC. The confidence parameters in this model are not a change in memory strength, but rather in response bias.

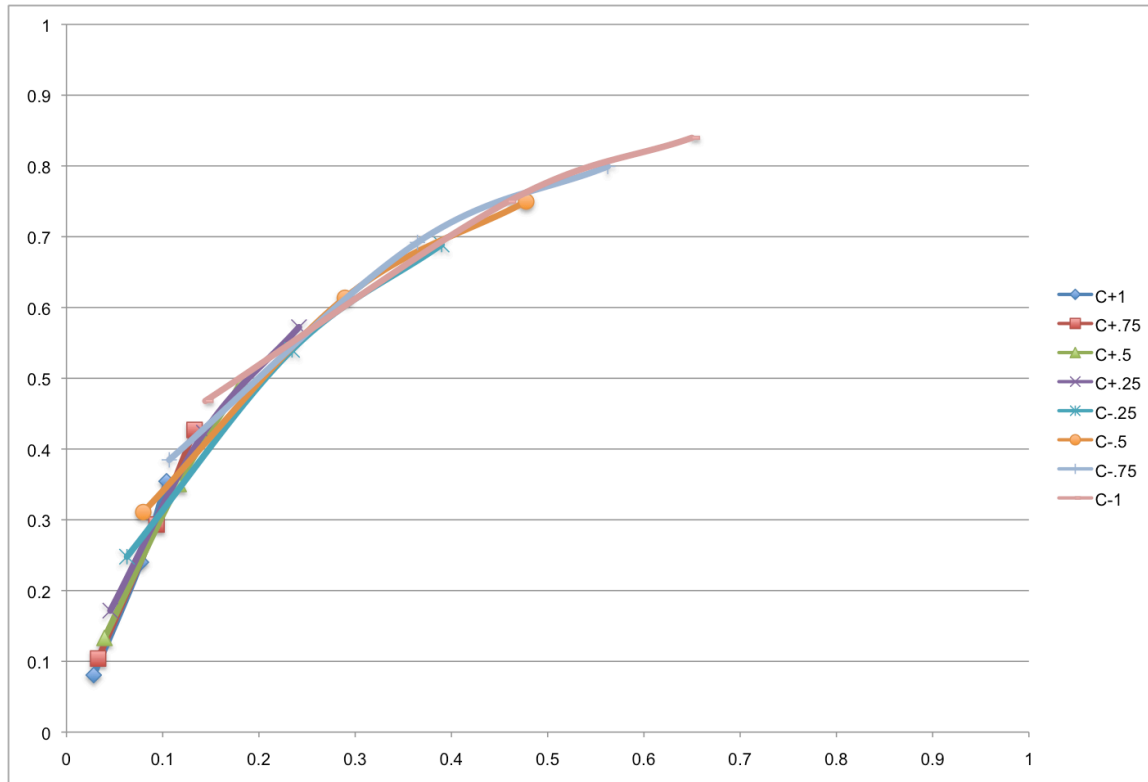
**Showup.** Increasing the confidence criteria parameters led to lower choosing rates, with lower hits, lower false alarms, and little change in  $d'$  as seen in Table 4.

Table 4.  
*Signal Detection Variables for Simulated and Real Showup Data Sets*

	Hit Rate	False Alarm Rate	d'	Beta
Collected Data	.83	.15	1.62	1.18
Ratcliff & Starns (2009)	.64	.31	0.85	1.06
+.25 to confidence	.57	.24	0.88	1.26
+.5 to confidence	.50	.18	0.89	1.51
+.75 to confidence	.43	.13	0.92	1.82
+1 to confidence	.35	.10	0.88	2.06
+.25 to decision	.42	.13	0.92	1.84
+.5 to decision	.23	.05	0.88	3.11
+.75 to decision	.11	.02	0.88	4.38
+1 to decision	.05	.02	0.53	2.72
-.25 to confidence	.69	.39	0.77	0.92
-.5 to confidence	.75	.48	0.73	0.80
-.75 to confidence	.80	.56	0.68	0.71
-1 to confidence	.84	.65	0.61	0.66
-.25 to decision	.80	.55	0.73	0.70
-.5 to decision	.91	.77	0.59	0.54
-.75 to decision	.89	.73	0.56	0.61
-1 to decision	.98	.96	0.36	0.49

*Note.* Confidence refers to a change in the confidence parameters in the model. Decision refers to a change in the decision parameter, with a – referring to a decrease in the yes decision bins and an increase in the no decision bins and vice versa.

Unlike the lineup simulation, the showup simulation showed higher betas. Again different from the lineup simulation, the showup simulation acted in a way more similar to the current understanding of ROC curves according to standard signal detection theory (see Figure 6).



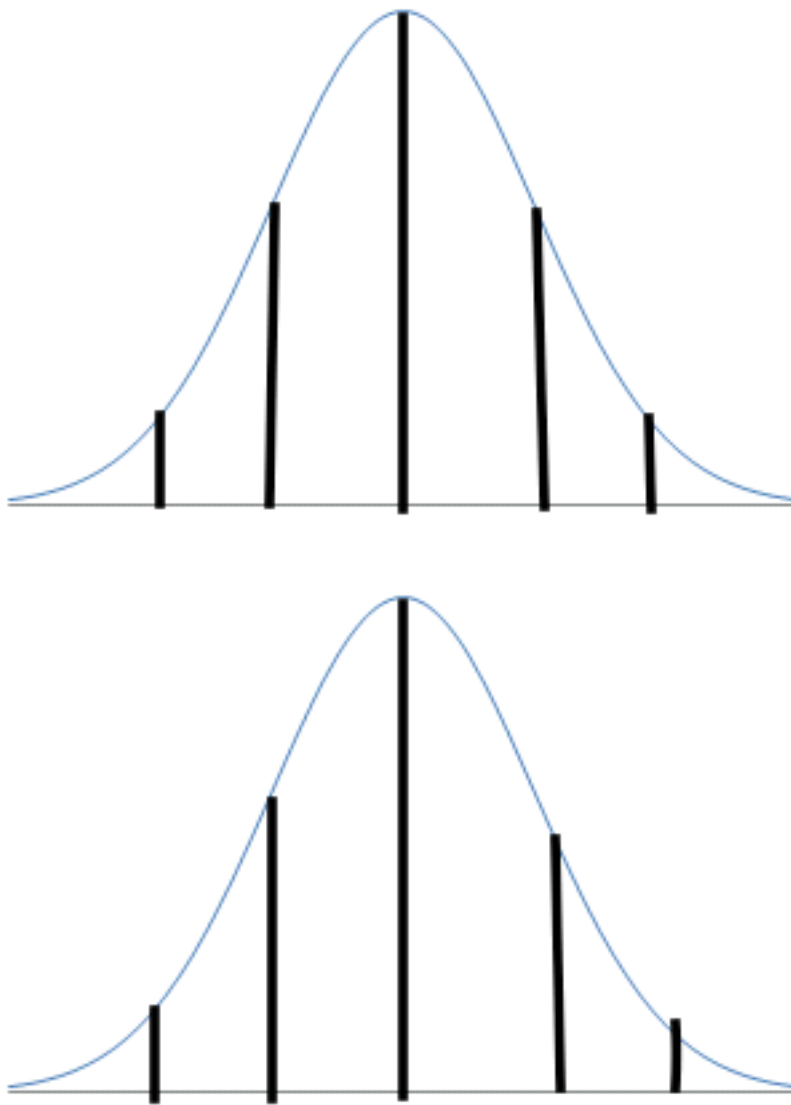
*Figure 6. ROC Curves of Simulated Showup Data with Confidence Criterion Manipulation*

The line shapes change little and there is a more even progression across the x and y axes. This even movement does not change the pAUC, which is expected by a change in response bias.

There is very little change in the shape to the curves, but there is movement up the y-axis with slight movements across the x-axis.

**Changes in decision criteria.** There are two different parameters in the RTCON model that simulate response criterion. The second is decision criterion. The basic premise of this parameter is that there is a zero sum game for the decision areas. As you increase one area (or

require more information to say yes), you must decrease evenly across the other areas (see Figure 7).



*Figure 7. Showing an Increase in the Fourth Decision Parameter*

As a reminder, the decision criteria were manipulated by adding across the yes criteria and subtracting across the no criteria or vice versa. Changes were made in .25 increments up to 1.

**Lineup.** Decreasing the decision criteria (requiring less information to say yes) causes very little change to the hit or false alarm rates, which in turn causes little change to  $d'$ , although there is a large change in  $d'$  from the .75 decrease to the 1 decrease. The biggest change in decreasing the decision criteria is a dramatic increase in beta. Increasing the decision criteria (requiring more information to say yes) causes a sharp decrease in hit rates but a comparatively small decrease in false alarms.  $D'$  changes consistently, but not overly largely. Beta moves very little. The ROC curve does undergo extreme changes as the parameters are manipulated (see Figure 8).

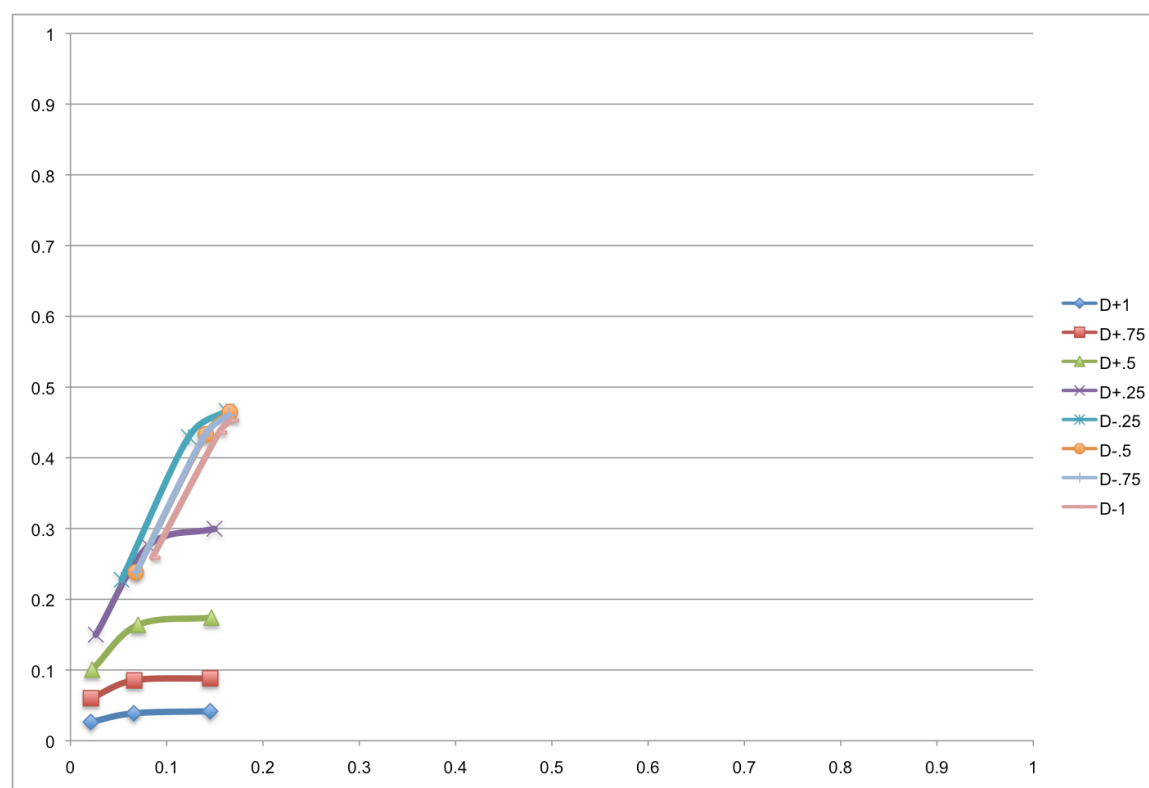
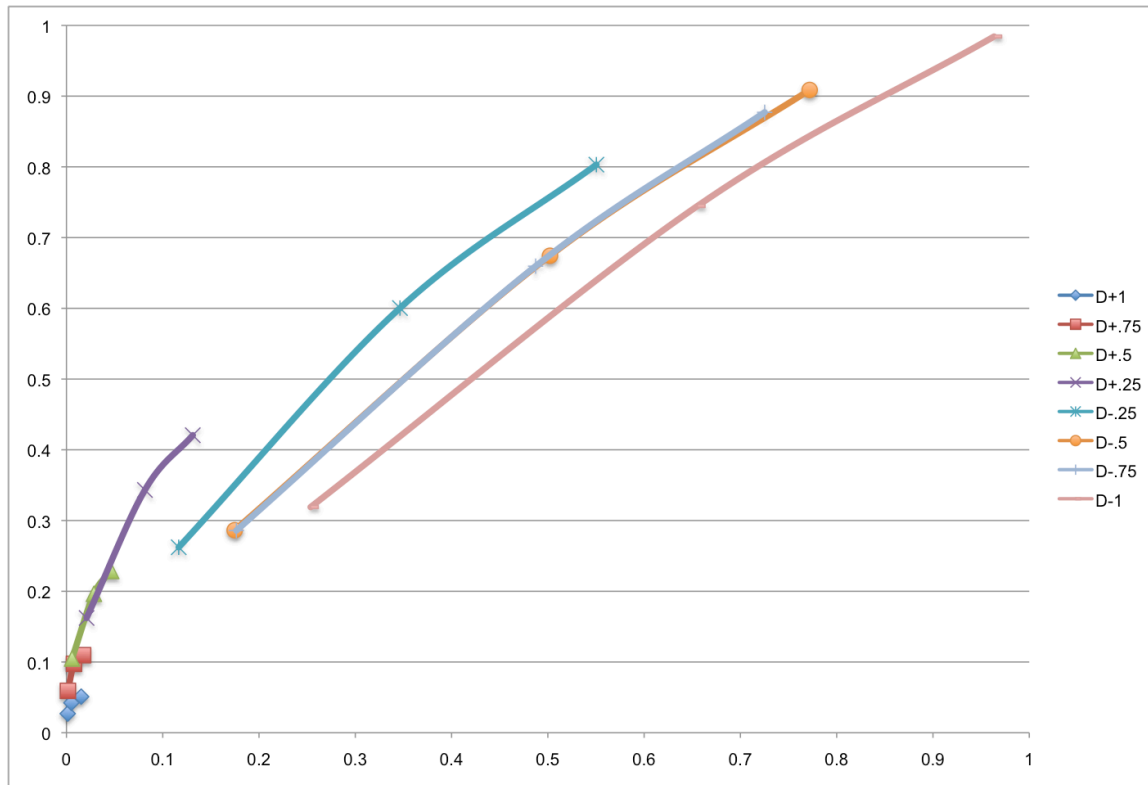


Figure 8. ROC Curves for Simulated Lineup with Decision Criteria Manipulation

When the most information is required to say yes, the curve is almost flat. As less and less information is needed, the line takes on a stronger curve and moves up the y-axis. Again this

increase in the pAUC as the decision criteria changes (by definition a change in response bias), is against what is theoretically proposed with ROCs.

**Showup.** Decreasing the decision criteria (requiring less information to say yes) causes little change to the hit rate (although this may be due to a slight ceiling effect). The false alarm rate does show marked increases as the decision criteria decreases. This inflation of false alarm does lead to some somewhat large decreases in  $d'$ . Beta moves around a good bit. Increasing the decision criteria (requiring more information to say yes) moves in the opposite way. Hit rates decrease dramatically. False alarm rates decrease somewhat (although this may be due to a floor effect). Beta moves around a bit in an inconsistent pattern.  $D'$  moves little between the first three increases of the decision criteria but drops dramatically from the third to the fourth change. The ROC curves change quite a bit through the manipulations of the decision criteria parameters (see Figure 9).



*Figure 9. ROC Curves for Simulated Showup with Decision Criteria Manipulation*

As you add the amount of information required to say yes, the line shrinks with slight movement across the axes. The bigger change is when the decision criteria is decreased. Two of the lines (d-.5 & d-.75) are virtually on top of each other, but the other decreases to decision criteria are very separate from those lines and there is a change in the pAUC. Again since decision criteria is not a memory strength parameter, but a response bias parameter, there should be no change in the pAUC.

### **Overall Comparisons Between Empirical and Simulated Data**

Again only qualitative comparisons can be made due to the restrictions of the project, but it is important for a preliminary comparison of the collected data and the simulated data to discuss the differences in ROC performance.

**Lineup.** This simulation of lineup data was woefully unable to simulate the pattern of hit rate, false alarm rate,  $d'$ , or beta that the collected data demonstrated. A specific difference from the model to the actual results was an extremely high false alarm rate that likely had a ceiling effect across the criteria manipulations. The false alarm rate was always several tenths higher than the hit rate, resulting in very low  $d'$ s and betas. None of the ROC curves are similar to the curves derived from the actual data either. The partial area under the curve (pAUC) for the actual data is much larger than that for the simulated data.

**Showup.** For the showup data, the model works a little bit better for the signal detection theory calculations, but is still lacking. False alarm rates do not ever have higher values than hit values. Because of this trend,  $d'$ s and betas are closer to the actual data as well. The actual participants still outperform the simulated participants pretty strongly with over twice the  $d'$  of the simulated participants at the hit rate level seen in the actual data. The ROC curve also shows a similar pattern to the lineup data with the pAUC being much greater in the real data, than can be seen with any simulated data.

## **General Discussion**

### **Behavioral Data**

In many ways the data for this project is similar to previous research. Lineups are more diagnostic than showups, which means it is better to be an innocent suspect in a lineup than a showup. This also means that lineups have stronger investigative evidentiary value than showups do. Showups do have more hits, but also more false alarms and a lower beta. One reason suggested for this is filler siphoning (Wells et al., 2015). Because lineups have fillers that siphon evidence, lineups are at an advantage over showups for lower false alarms. This hypothesis may be backed up by the significant difference in response time between presentation



types. People who were shown a showup responded faster than participants who were shown a lineup. Because the fillers are siphoning some of the possible choices during identification, it would take longer for any choice to reach a decision threshold.

One surprise from the behavioral data was that instruction bias and presentation type interacted causing an increase in confidence choice for lineups going from biased to unbiased instructions. Typically, unbiased instructions make participants less sure in their choice, possibly because it is highlighting potential reasonable doubt. Showups do adhere to the typical confidence pattern. It is possible that because the participants were given an admonishment it forced them to have a higher threshold for decision, which made them more confident in their choice as more evidence was collected under any choice made.

### **ROC Debate**

There has been recent controversy in the cognitive sphere over the use of ROC curves to examine confidence and accuracy (e.g. Gronlund, Wixted, & Mickes, 2014; Lampinen, 2016, Wells et al., 2015). Proponents say that ROC curves provide a better measure of lineup performance because ROC curves provide an indication of memory strength that is independent of response bias. According to the standard account, the shape of the ROC – i.e., the degree to which it is bowed upwards towards the top left corner – provides an indication of memory strength, whereas the location of an individual response along the X-Axis provides an independent measure of response bias. This interpretation makes sense in terms of standard unequal variance signal detection models, but falls apart under stochastic models such as Ratcliff and Starns (2009) RTCON and Van Zandt, Colonius, and Proctor's (2000) Poisson counter confidence model.

The current findings also call into question recent interpretations of the practical utility of ROC analyses. For instance, Wixted et al. (under review) have argued that when one partial ROC curve dominates another partial ROC curve, even if the two curves are truncated at different points along the X-Axis, that it should always be possible to equate the two conditions in terms of false alarm rates by simply manipulating response criteria via instruction. The presented data shows that this is simply not the case. Manipulating response criteria via instruction appears not only to change the false alarm rate but can also impact the hit rate, changing the shape of the ROC curve. The instruction bias manipulation is looked at in the literature as a response bias change. Using the ROC curves in Figure 1, the unbiased ROC curves dominate the biased ROC curves for both lineup and showups. This would mean in the traditional interpretation, that instruction bias is somehow affecting memory strength. The theoretical basis for instruction bias completely refutes this interpretation. This means that in the collected data, our actual participants were answering in ways that are completely contrary to how ROC curves are being used in the eyewitness literature today.

The simulated data presented here also speak to how appropriate it is to use ROC curves for this type of data. In some instances, there is a confirmation of the typical use and in some instances a refutation of these beliefs. Specifically, the ROC curve of the simulated data for showups largely validate the current interpretation of ROC curves. The manipulation of  $\mu$  shows an increase in memory strength causing a shift up the y-axis (see Figure 3). The manipulation of the confidence parameter (something that would mimic a change in response bias) shows a clear shift along the x-axis and no significant change in the shape of the curve (see Figure 6).

For lineup data there is a very different pattern shown. Both the memory strength manipulation and the confidence manipulation cause a shift along the y-axis (see Figure 2 and 5, respectively). While the memory strength manipulation is in line with current thinking of ROC behavior, the pattern in the confidence manipulation directly contradicts current thinking with a migration along the y-axis but no movement along the x. If anything, the ROC curves for confidence criterion, show a more clear migration up the y-axis than the memory strength manipulation. The ROC curves for the decision criteria manipulations also contradict current thinking. The lineup simulations show a pattern that replicates the migration along the y-axis that was seen with the confidence manipulation, but also include changes in the shapes of the ROC curves, another element that should be restricted to memory strength manipulations (see Figure 8). The showup data also somewhat refutes this thinking. While there is movement along the x-axis in the showup decision criteria simulations, there are also changes to the shape of the curves and movement along the y-axis (see Figure 9).

The decision criterion and the confidence criterion, theoretically, can only be changes in response bias. As such the simulations provided here show a clear refutation to the common thinking of ROC curves. It also provides clear evidence that they may not be the appropriate tool in every circumstance to differentiate between methods in the eyewitness literature.

### **RTCON Simulations**

The RTCON model may be applicable for some stimuli, but seems woefully inept for the stimuli used in this study. None of the simulated data approached the collected data with regards to accuracy. Even at the most liberal criterion shifts for lineups, the average hit rate of the collected data is higher (.69) than that of the simulated data (.57 at -1 to the confidence criterion). The simulated data fares a little better in the showup situation. While there are instances where

the hit rates outperform the collected data, they are accompanied by such high false alarm rates that they are diagnostically useless. In both showup and lineup instances,  $d'$  is much higher for the collected data than the simulated data.

These results suggest that there is something fundamentally different in the current experiment compared to the experiment used by Ratcliff and Starns (2009). This experiment exposed participants to the face they would have to identify for eight seconds (a time span used in other eyewitness literature) in a video. The experiment Ratcliff and Starns used word pair stimuli that were presented for 1.8 seconds. The current experiment also tested memory immediately following the video. The RTCON experiment went through the entire study phase (36 word pairs) before beginning the test phase. These differences were largely due to applications of the current research. As it was the intention to apply the RTCON model to eyewitness literature, it was more important to be as ecologically valid as possible, rather than attempting a direct replication.

One aspect that has not been fully discussed is the role that the instruction manipulation had in relation to the RTCON model. There is not a fundamental difference in simulating biased compared to unbiased data as there was for the lineups compared to showups. Rather to examine the effectiveness of using the RTCON model on this manipulation, it is better to look at the individual cells of the collected data compared to liberal or conservative criteria. This would mean examining biased instruction data in relation to liberal criteria (shifting the confidence criterion distribution to the right or decreasing the level of evidence needed to say “yes” in the decision criterion distribution) and unbiased instruction data in relation to conservative criteria (shifting the confidence criterion distribution to the left or decreasing the level of evidence needed to say “no” in the decision criterion distribution).

Admittedly when examining these statistics, it is important to again note the strange behavior of the participants with regard to the instruction manipulation. Although biased instructions typically yield higher hit rates and higher false alarms, this is not what is found in this data. A better  $d'$  does occur for unbiased lineup presentations compared to biased lineup presentations, which does fit with the typical data. Biased showup presentations had a better  $d'$  than unbiased showup presentations. Although the raw numbers for lineup simulations are not on par with the actual data, a similar trend to the  $d'$ 's when moving from conservative to liberal response bias is shown. Or in other words,  $d'$  increases as response bias becomes more conservative (regardless of which criterion was manipulated). The showup data did not agree though. Again  $d'$  increases as response bias becomes more conservative. This finding though is in contrast to what was found in the actual data.

These comparisons in general suggest that the RTCON model is not necessarily relevant to eyewitness identifications. The simulated raw measures were not close to those found in the actual data. Additionally, the pattern of movement in the measures sometimes differed from the collected data. When one aspect matched the collected data (a similar hit rate), it would drastically differ in another (completely different false alarm rate). This might suggest that any similarities observed could be due to chance as there does not seem to be a strong theoretical background to explain these differences.

### **Other Confidence Theories**

So if RTCON is not acceptable for eyewitness identification, what are some of the other possibilities to explain the complex process of choice and confidence when choosing someone as a suspect? There currently does not seem to be a theory that is consistently accepted among researchers in this area. Instead there are a few different theories that have been debated.

The first is based on signal detection theory. It says that confidence is matched with the level of memory strength for the target (Wixted & Mickes, 2014). If it is a strong match, a high level of confidence will be chosen. If it is a weak match, a low level of confidence will be chosen. While this works in dichotomous, yes/no scenarios, it is less useful in a lineup situation. If a witness used this method of confidence choice, they are ignoring the foils in a lineup situation. This is largely against the believed cognitive processes that go into identifying a suspect and many researchers disagree with its applicability to a lineup procedure.

A slightly more nuanced theory does take into account possible alternatives. Recently, Horry and Brewer (2016) applied sequential sampling model elements to confidence choices following decisions. Specifically, they said that confidence choice is based on the relative success of the decision over the alternatives. In other words, if the choice you make has a lot of evidence accumulated for it compared to the alternatives, it is going to be a high confidence choice. If the choice you make crosses the decision boundary threshold, but another choice is close to the boundary, it is going to be a low confidence choice. This theory better represents the complex cognitive processes that goes into a multiple choice decision rather than treating each choice individually like in the signal detection theory discussed above.

## **Conclusions**

Due to unforeseen circumstances, the data simulation originally proposed for this project had to be worked around. This leaves the data simulation discussed here in a bit of limbo. It is possible that with a different method of simulation, a different pattern of results may emerge. Given the simulations that were done, however, it seems like RTCON is not appropriate for an eyewitness paradigm. Instead, it would serve researchers to examine different theories of decision and confidence to try and achieve a better understanding of the underlying mechanisms

that affect both. Ideally, a theory seeking to explain these aspects should not only work in a two choice showup situation, but should be just as applicable in a multiple choice lineup situation. The cognitive processes are slightly different as showups require an absolute judgment of recognition or one that relies solely on the match between stimulus and memory strength, but lineups can use a combination of both absolute and relative judgments when making a choice. As such it is important to have some understanding of the underlying mechanisms for both procedures, especially because sometimes a witness may participate in both with the same suspect during the course of an investigation.

The behavioral data also showed that the participants in this study did better than most participants do on eyewitness tasks. A possible suggestion for this is that the participants in this study completed eight trials each. Most eyewitness research uses one single crime with a single perpetrator and a single choice is asked for. It is possible that participants learned the format of the crimes shown and the presentation of the suspect. This could give them a memory advantage or at least inform the participants of a possible strategy for the later trials. The number of trials was required in order to examine the model in question, but it is important to note this possible confound in the behavioral data.

Another key, though unexpected, finding from this research directly contradicts the signal detection theories that underlie ROC curves. Specifically, the data simulation was able to demonstrate that ROC curves can be more subject to changes in response bias than previously believed. Rather than just a change along the false alarm or x-axis, changes in the confidence criterion for lineups showed a direct change along the hit or y-axis. According to signal detection theory, this should only be possible if the confidence criterion affected memory

strength, something that is impossible in the RTCON model. This contributes to the literature suggesting that ROC curves may not be as useful in the eyewitness realm as previously believed.

While this research did not end in the place expected at the beginning of the process, it has contributed some interesting and important findings to the overall literature. It also keeps the door open for another theory of eyewitness decision and confidence relationships. The only way to interpret the data in this study, though, is that the collected data and simulated data exist in theoretically separate environments.



## References

- Bradley v. State, 264 SD.E.2d 332, (1980).
- Brewer, N. & Wells, G. L. (2006). The confidence-accuracy relationship
- Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior*, 7, 19-30.
- Carlson, C. A. (2011). Influence of a perpetrator's distinctive facial feature on eyewitness identification from simultaneous versus sequential lineups. *Applied Psychology in Criminal Justice*, 7, 77-92.
- Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, 29, 575-604.
- Douglass, A. B. & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology*, 18, 152-167.
- Flowe, H. D., & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior*, 31, 33-52.
- Foster, R. A., Libkuman, T. M., Schooler, J. W., & Loftus, E. F. (1994). Consequentiality and eyewitness person identification. *Applied Cognitive Psychology*, 8, 107-121.
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operator characteristic analysis. *Current Directions in Psychological Science*, 23, 3-10.
- Horry, R., & Brewer, N. (2016). How target-lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General*. Advanced Online Publication.
- Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory and Cognition*, 5, 21-33.
- Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior*, 33, 194-212.
- Levi, A. M. (1995). The police lineup: Basic weaknesses, radical solutions. *Criminal Justice and Behavior*, 22, 347-372.
- Lindsay, R. C. L., Kalmet, N., Leung, J., Bertrand, M. I., Sauer, J. D., & Sauerland, M. (2013). Confidence and accuracy of lineup selections and rejections: Postdicting

- rejection accuracy with confidence. *Journal of Applied Research in Memory and Cognition*, 2, 179-184.
- Lindsay, R. C. & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, 70, 556-564.
- Loftus, E. F. (1974) Reconstructing memory: The incredible eyewitness. *Psychology Today*, 8, 116-119.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: lineup instructions and the absence of the offender. *Journal of Applied Psychology*, 66, 482-489.
- Manson v. Braitwaite, 432 US 98 (1977)
- McClure, K. A., Myers, J. J., & Keefauver, K. M. (2013). Witness vetting: What determines detectives' perceptions of witness credibility?. *Journal of Investigative Psychology and Offender Profiling*, 10, 250-267.
- Mickes, L., Flowe, H. D., & Wixted, J. T., (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, 18, 361-376.
- National Institute of Justice. (2013). A national survey of eyewitness identification procedures in law enforcement agencies. Washington, DC: Police Executive Research Forum
- Navon, D. (1992). Selection of foils by similarity to the suspect is likely to misfire. *Law and Human Behavior*, 16, 575-593.
- (n.d.) *Eyewitness Misidentification*. Retrieved from <http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php>
- (n.d.) *Ronald Cotton*. Retrieved from [http://www.innocenceproject.org/Content/Ronald\\_Cotton.php](http://www.innocenceproject.org/Content/Ronald_Cotton.php)
- Neil v. Biggers, 409 US 188, (1972)
- Odinot, G., Wolters, G., & van Koppen, P. J. (2009) Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months. *Law and Human Behavior*, 33, 506-514.
- Paley, B., & Geiselman, R. E. (1989). The effects of alternative photospread instructions on suspect identification performance. *American Journal of Forensic Psychology*, 7, 3-13.

- Pozzulo, J. D., & Dempsey, J. (2006). Biased lineup instructions: Examining the effect of pressure on children's and adults' eyewitness identification accuracy. *Journal of Applied Social Psychology*, 36, 1381-1394.
- Quinlivan, D. S., Neuschatz, J. S., Cutler, B. L., Wells, G. L., McClung, J., & Harker, D. L. (2012). Do pre-admonition suggestions moderate the effect of unbiased lineup instructions, *Legal and Criminological Psychology*, 17, 165-176.
- Rahaim, G. L., & Brodsky, S. L. (1982). Empirical evidence versus common sense: Juror and lawyer knowledge of eyewitness accuracy. *Law & Psychology Review*, 7, 1-15.
- Rankin, B. (2008, December 25). Success in improving Georgia's eyewitness identification procedures. *The Atlanta Journal-Constitution*. Retrieved from: [https://www.schr.org/content/success\\_in\\_improving\\_georgias\\_eyewitness\\_identification\\_procedures](https://www.schr.org/content/success_in_improving_georgias_eyewitness_identification_procedures)
- Ratcliff, R., McKoon, G., & Tindall, M. H. (1994). Empirically generality of data from recognition memory receiver-operating characteristic functions and implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 763-785.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116, 59-83.
- 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, 697-719.
- Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior*, 36, 394-403.
- Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification. *Journal of Applied Psychology*, 89, 334-346.
- Smith, A. M., Bertrand, M., Lindsay, R. C. L., Kalmet, N., Grossman, D., Provenzano, D. (2014). The impact of multiple showups on eyewitness decision-making and innocence risk. *Journal of Experimental Psychology: Applied*, 20, 247-259.
- Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed?. *Journal of Applied Psychology*, 85, 542-550.
- Smith, S. M., Lindsay, R. C. L., Pryke, S., Dysart, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation?. *Psychology, Public Policy, and Law*, 7, 153-169.

- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22-33.
- Stebly, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-297.
- Stebly, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003) Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*, 523-540.
- Stinson, V., Devenport, J. L., Cutler, B. L., & Kravitz, D. A. (1996). How effective is the presence-of-counsel safeguard? Attorney perceptions of suggestiveness, fairness, and correctability of biased lineup procedures. *Journal of Applied Psychology, 81*, 64-75.
- Stovall v. Denno, 388 US 293, (1967)
- Tetterton, V. S. & Warren, A. R. (2005). Using witness confidence can impair the ability to detect deception. *Criminal Justice and Behavior, 32*, 433-451.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*, 208-256.
- Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64*, 440-448.
- Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition, 4*, 313-317.
- Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understand and applied purposes. *Journal of Applied Research in Memory and Cognition, 4*, 324-328.
- Wells, G. L., Steblay, N. K., & Dysart, J. E. (2014). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*. Advance online publication.
- Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition, 4*, 8-14.
- Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121*, 262-276.

- Wixted, J. T., & Mickes, L. (2015). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, 4, 329-334.
- Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 113, 304-309.
- Wolgalter, M. S., Marwitz, D. B., & Leonard, D. C. (1992). Suggestiveness in photospread lineups: Similarity induces distinctiveness. *Applied Cognitive Psychology*, 6, 443-453.

## Appendix A

## Letter of Research Compliance Approval



Office of Research Compliance  
Institutional Review Board

October 3, 2014

## MEMORANDUM

TO: Brittany Race  
James Lampinen

FROM: Ro Windwalker  
IRB Coordinator

RE: New Protocol Approval

IRB Protocol #: 14-09-137

Protocol Title: *Confidence in Eyewitness Identification: A Diffusion Model Approach*

Review Type: ☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB

Approved Project Period: Start Date: 10/03/2014 Expiration Date: 10/02/2015

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (<http://vpred.uark.edu/210.php>). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

**This protocol has been approved for 2,000 participants.** If you wish to make any modifications in the approved protocol, including enrolling more than this number, you must seek approval prior to implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 210 Administration Building, 5-2208, or [irb@uark.edu](mailto:irb@uark.edu).